
# TOWARD VALIDATING VERTICAL AGGREGATION IN OBJECT ORIENTED SIMULATION

THESIS

George W. Johnson Jr., First Lieutenant, USAF

AFIT/GOR/ENS/00M-17

**DEPARTMENT OF THE AIR FORCE**

**AIR UNIVERSITY**

*AIR FORCE INSTITUTE OF TECHNOLOGY*



**Wright-Patterson Air Force Base, Ohio**

| REPORT DOCUMENTATION PAGE | Form Approved OMB No. 0704-0188 |
|---|---|



| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From - To) |
|---|---|---|
| 14-03-2000 | Master's Thesis | Oct 1999-Mar 2000 |

**4. TITLE AND SUBTITLE**

Toward Validating Vertical Aggregation In Object Oriented Simulation

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

George W. Johnson Jr., First Lieutenant, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 P Street, Building 640
WPAFB OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AFIT/GOR/ENS/00M-17

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Maj. Wallace Langbehn
HQ USAF/XOC
1480 Air Force Pentagon
Washington D.C. 20330-1480
DSN: 425-5065

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Recent interest in studying the nonlinear effects of intelligence, surveillance, and reconnaissance in combat models has prompted researchers to employ vertical aggregation in object-oriented simulations. Traditional horizontal aggregation falls short for its inability to provide accurate means for nonlinear functions. Averaging a group of objects that exhibit nonlinear behavior provides a linear approximation to the mean, which is not necessarily the expected value of the underlying nonlinear function. Vertical aggregation explicitly models individual objects, thus preserving their nonlinear behaviors. In this research, a validation procedure is derived to study the aptness of vertical aggregation methods. Validation is carried out by comparison with a control, considered model truth, since it contains no vertical aggregation. Response surfaces are mapped for the control and the hypothesized model. Family confidence intervals are used to test the hypothesis that the difference between the two is zero. An illustrative example is presented using a homogeneous combat scenario embellished with experimental factors. Metamodels are derived using the method of least squares and validated prior to drawing inferences. Simultaneous inferences are drawn between the ith regression coefficient of two models. The results suggest fascinating avenues for further study.

**15. SUBJECT TERMS**

Nonlinear Effects, Intelligence, Surveillance, Reconnaissance, Combat Model, Vertical Aggregation, Object Oriented, Object Scaling, Validation, Comparison With Control, Response Surface, Metamodel

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Maj. W. Paul Murdock |
| Unclassified | Unclassified | Unclassified | UL | 73 | 19b. TELEPHONE NUMBER (Include area code) DSN: 785-6565 x4339 |

**The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U. S. Government.**

AFIT/GOR/ENS/00M-17

# TOWARD VALIDATING VERTICAL AGGREGATION IN OBJECT ORIENTED SIMULATION

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

George W. Johnson Jr.

First Lieutenant, USAF

March 2000

AFIT/GOR/ENS/00M-17

# TOWARD VALIDATING VERTICAL AGGREGATION IN OBJECT ORIENTED SIMULATION

George W. Johnson Jr.

First Lieutenant, USAF

Approved:

| | |
|---|---|
| W. Paul Murdock Jr., Maj., USAF (Advisor) | 15 MAR 00 date |
| John S. Crown, Maj., USAF (Reader) | 15 Mar 00 date |
| John O. Miller, Lt. Col., USAF (Reader) | 15 Mar 00 date |

**Acknowledgements**

Thank you Lisa, Timothy, and Amanda. You are not only my inspiration, but you truly define me. For all of my days, past and yet here, I strive to make your world a better place.

Thanks to my many advisors; academic and professional. Your advice and insight binds these pages together.

George W. Johnson Jr.

## Table of Contents

## List of Figures

## List of Tables

**Abstract**

Recent interest in studying the nonlinear effects of intelligence, surveillance, and reconnaissance in combat models has prompted researchers to employ vertical aggregation in object-oriented simulations. Traditional horizontal aggregation falls short for its inability to provide accurate means for nonlinear functions. Averaging a group of objects that exhibit *nonlinear* behavior provides a *linear* approximation to the mean, which is not necessarily the expected value of the underlying nonlinear function. Vertical aggregation explicitly models individual objects, thus preserving their nonlinear behaviors.

In this research, a validation procedure is derived to study the aptness of vertical aggregation methods. Validation is carried out by comparison with a control, considered model truth, since it contains no vertical aggregation. Response surfaces are mapped for the control and the hypothesized model. Family confidence intervals are used to test the hypothesis that the difference between the two is zero.

An illustrative example is presented using a homogeneous combat scenario embellished with experimental factors. Metamodels are derived using the method of least squares and validated prior to drawing inferences. Simultaneous inferences are drawn between the $i^{th}$ regression coefficient of two models. The results suggest fascinating avenues for further study.

# TOWARD VALIDATING VERTICAL AGGREGATION IN OBJECT ORIENTED SIMULATION

## I. Problem

### Introduction

Models represent reality to varied degrees of accuracy. Simplifying assumptions are essential to properly scope an analysis project. The intent of simplification is to improve the efficiency of the analysis with minimal degradation in output fidelity. Thus, a model with the most detail imaginable will be the least efficient and most accurate – given that those details are themselves accurate. Then as the model moves from more to less detail, the efficiency should increase as the fidelity of the output decreases. Good analysts will find a balance that best suits the problem being studied. The trick is to cut the details that have the least impact on the measures being studied – a difficult task since their impact is not clear in the first place.

Such difficulty does not vanish when the analyst chooses the System Effectiveness Analysis Simulation (SEAS). However, with SEAS the analyst has a tool with which to cut the detail while reportedly maintaining the utmost integrity in the measures being studied. If used correctly, theory suggests that the analyst may significantly reduce the detail of the model while maintaining essential complex

interactions between the model's objects. Thus, output that is largely dependent on such interactions will be a good representation of the more complicated model's output.

*The Vertical Slice Methodology*

The vertical slice methodology is an artistic analytical method for reducing a complicated model to its simplest form. Reduction is accomplished by *object scaling* – deleting a large number of redundant objects in the simulation. The model builder creates a simulation using a representative quantity of each object class. How different objects are scaled is largely dependent on the scenario and problem under consideration. It is not necessary – or likely – that all objects be scaled by the same factor. It is, however, necessary to maintain proportionality among objects with respect to their *dependencies* and *interactions*. To that end, the analyst must understand how the objects' threads are manifested in the simulation. When properly understood, this manifestation may be adequately captured with much less redundancy.

> In setting down some guidelines for constructing a "vertical slice" scenario...the most important objective is to capture in a balanced manner the significant dependencies and interactions among forces, sensors, weapons and targets....[T]he analyst constructs a "vertical slice" by removing the majority of objects (e.g., 4 out of 5) across all units on the battlefield. The objective of this approach is to allow non-linear interactions among units. (SMC/XR:1).

Applying the vertical slice methodology, the analyst may realize an 80% to 90% reduction in the number of entities or objects simulated. Interactions between objects may be reduced by perhaps 98% or more. Thus, the vertical slice methodology allows the analyst to drastically reduce the processing time of a simulation. Theory postulates

when slices are properly built, great run-time reduction may be realized without altering those measures of effectiveness central to the analysis.

*Object Scaling is Simple*

Consider a simple system to illustrate the problem (Figure 3). The goal is to shore up a dyke in the shortest time possible. A group of 60 people is assembled in six lines of ten people each. Suppose 6,000 sandbags are passed through the lines to be placed on the dyke. The period of interest is the time between the first sandbag entering and the last sandbag leaving the system.

One method for constructing a computer simulation of this problem will model each of the 60 individuals as a separate object. The resulting full-scale object model may give the most *realistic* simulation. Unfortunately, it is also the most processor-intensive simulation.

An alternative method will group objects of similar type and performance into separate classes. Since all objects in this example belong to a single class, each of the six lines should have similar – if not identical – performance. If each line operates independently, then one line may be modeled instead of six. By reducing the system to one line of ten people with 1,000 sandbags – objects and input reduced by one-sixth – the simulation will yield similar output – the average time to completion should not be significantly different. Unfortunately, the central limit theorem implies that scaling the model down will increase the variance of the average time to completion from the full-scale model since there are now fewer independent lines operating during each trial. The

degree of variance increase will be a function of the input parameters and their statistical distributions (Wackerly et al., 1996:328).

Object scaling is simple to apply to independent objects in a simple system. The scaled model is easier to build and verify. The simulation executes much faster, allowing the analyst to quickly remedy scaling induced variance increases through increased repetition. The resulting scaled model output gives the analyst essentially the same information as the full-scale model. Now, when faced with a system that is not so simple, how does one enjoy these benefits without losing the very complexity being analyzed?

*The OODA Loop Holds the Key*

To understand how to scale complex systems the analyst must understand the interactions between the system's objects. Each object has a decision cycle – often referred to as an OODA loop – during which it observes, orients, decides, and acts. The cycle is illustrated in Figure 1. The interactions and dependencies in a complex system are defined by the interaction of these loops (Tighe, 1999:6-20).

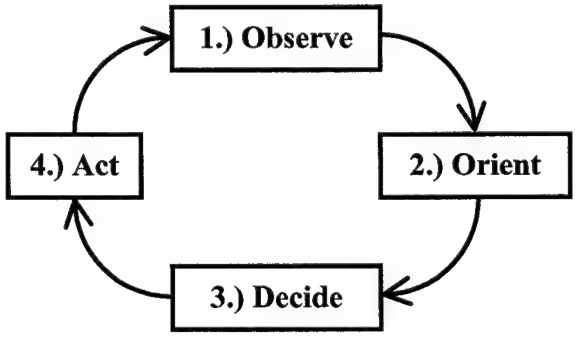


**Figure 1 The OODA Loop**

As Figure 2 shows, the simplest interaction occurs directly between two isolated objects. An object is influenced by – and has influence on – a second object. The actions of each feed the observations – and ultimately the actions – of the other. The two OODA loops are connected, though there need not be a two-way connection. Any one-on-one duel may exhibit such interaction of decision cycles.

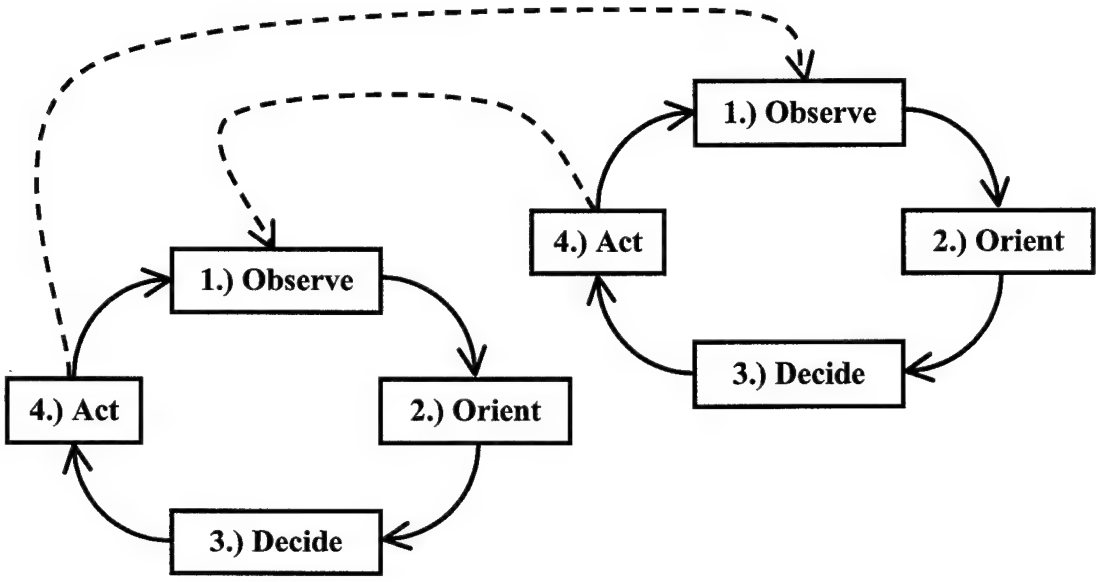


**Figure 2 OODA Loops Interact**

Figure 3 depicts the interactions in the dyke example. Assume a simple system in which interactions are confined to within each line. Each object has at most three objects to observe: itself and those adjacent to it in line – as seen in Figure 4. An object orients itself by comparing its observations to its rules. It then decides among three actions: catch, wait, or throw. The lack of interaction between OODA loops from one line to the

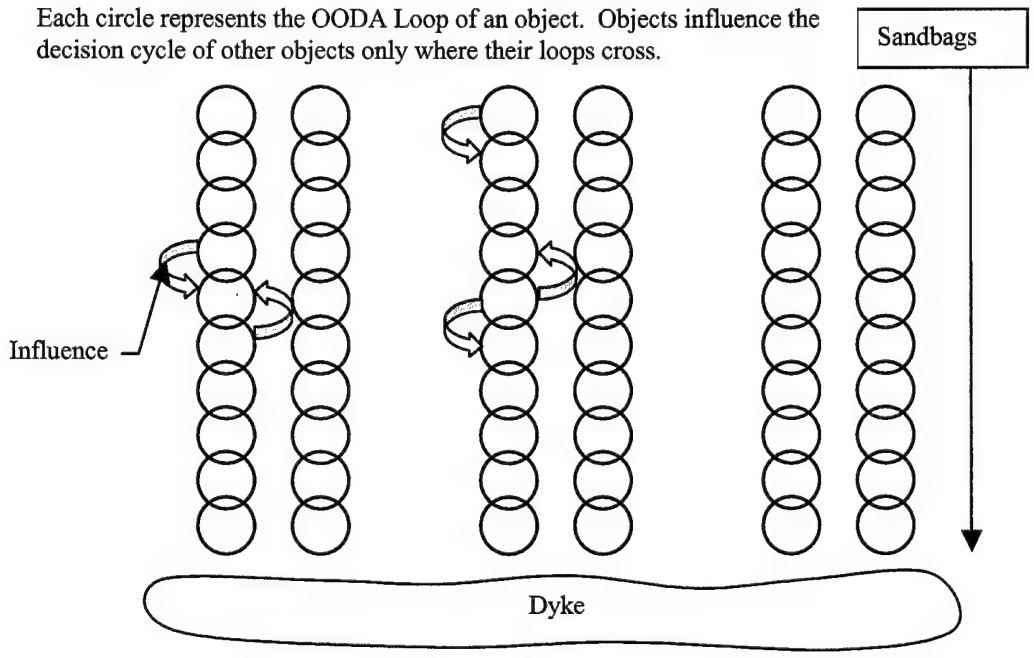


**Figure 3 Example Interactions**

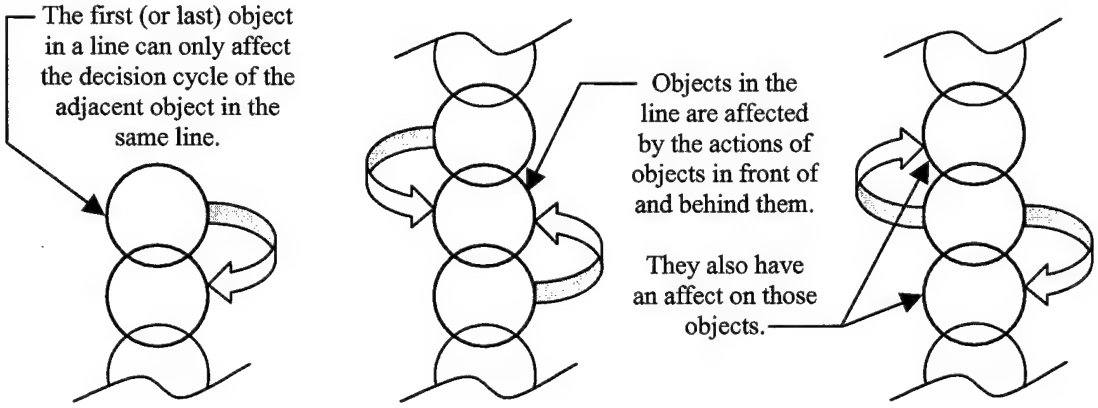


**Figure 4 Exploded View of Interactions**

next implies that the lines are independent. Thus, scaling to one line is justified. (Of course, the closed form solution for a single line can be easily found. The example is intended only to introduce the reader to the subject.)

*Building a "Slice" is Difficult*

A vertical slice is not interesting in simulations with no complex dependencies or interactions. In such a case – as in the above example – the vertical slice methodology is merely object scaling. A true vertical slice scenario will not be needed until there is a need to study a complex – not necessarily adaptive – system. Such complexity can be illustrated by adding rules to the dyke example.

Suppose that each person in the dyke example is subject to infrequent random events that cause incapacitation for some period of time. Allow the sandbags to be diverted to surrounding lines. Suppose that each person may pass 90° to either side or 45° forward to circumvent an incapacitated person (see Figure 5) and prohibit the sandbags from being passed further than one line from their original line.

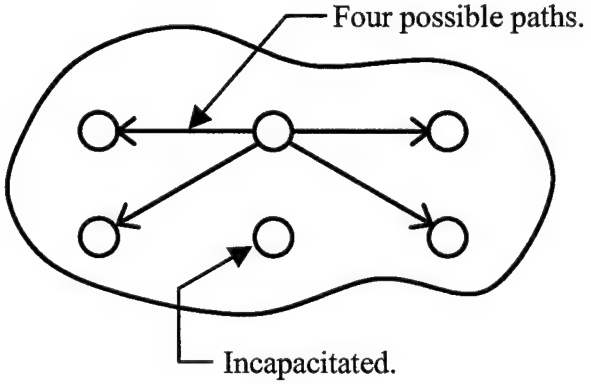


**Figure 5 Possible Sandbag Paths Around a Blockage**

The system thus changes from a simple one with 108 possible interactions to a complex one with 760 possible interactions. In the complex system, 32 objects have 16 interactions each. Eight objects in the same system have only six interactions each. By

contrast, the simple system had 12 objects with two interactions each, and 48 objects with four interactions each. Each line in the simple system is independent and identical. Notice the new complex structure of the system in Figure 6. The lines are now highly dependent and not identical. Lines one and six have probabilistically identical paths.

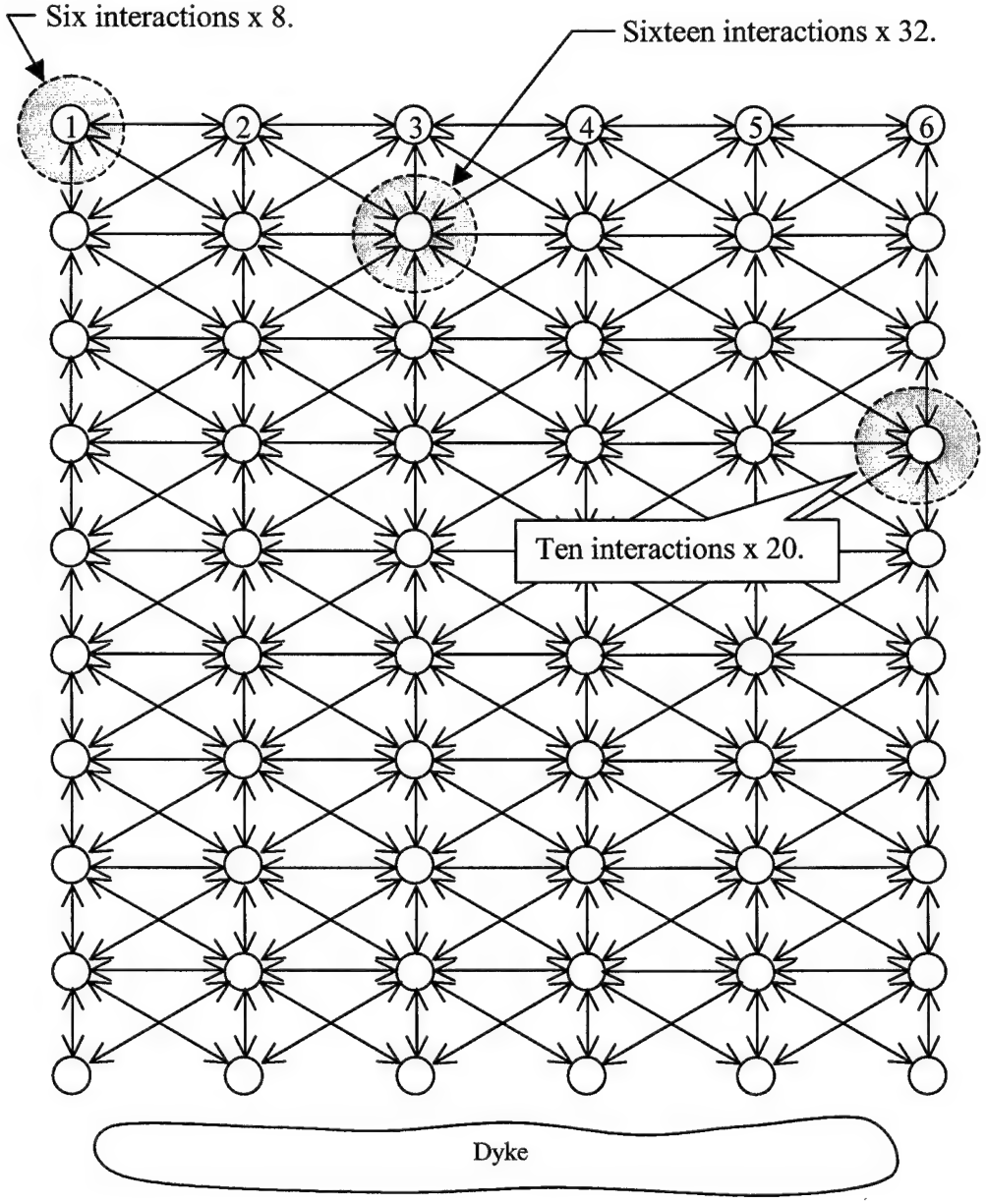


**Figure 6 Interactions and Dependencies in the Complex System**

A sandbag placed into the system in line one will have the same expected number of passes to get through if placed in line six – they belong to the same class. Lines two, three, four and five belong to a second class. A vertical slice of the dyke example would have two classes – outer lines and inner lines. One might construct a half-scale scenario with two inner lines and one outer. A second outer line – a dummy – could be used to preserve the interactions that define an inner line.

Further the illustration of complexity by allowing each person to see and judge – randomly and imperfectly – the actions and status of the surrounding people in a two or three-person radius. Add a supervisor to direct and an assistant for surveillance and reporting. Finally, allow the entities to adapt. The closer the system gets to reality the more complex it becomes and the more difficult it is to scale.

The example illustrates the heuristic nature of the vertical slice methodology. Analysts face problems – even in such a simplistic example: 1. It is difficult to scale objects that interact on a basis other than one-to-one. Object scaling becomes more difficult as the number of interactions and disparity among them increases. 2. Methods for scaling behaviors – such as perception and judgement – are not straightforward. The method will depend upon how the object is employed.

*The SEAS Approach*

The SEAS Analyst Manual reports that SEAS is not intended for modeling all objects on the battlefield. "Rather, we take a representative 'vertical slice'…from the… commander down to the air or ground platform level." The manual argues that by representing "…the range of heterogeneous interactions on the main campaign in the

scaled-down 'vertical-slice,' then the full battle is primarily the sum of homogeneous copies of that slice." With this approach, it is important that the slices be independent (SMC/XR:3). Any departure from independence may seriously corrupt the analysis. Thus, each slice must represent every inter-dependent complexity the analyst wishes to study. Complex issues modeled in one slice must be separable from those in all other slices.

### Topic

The *Vertical Slice Methodology* used in the System Effectiveness Analysis Simulation (SEAS) is not well known. Little guidance exists to relate theory to application. Where guidance does exist, it has not been validated by rigorous analytical methods.

This study explores the effects of object scaling in intelligence surveillance and reconnaissance (ISR) modeling. It investigates a specific object scaling scheme with respect to its suitability for preserving the characteristics of the region of interest. Special attention is given to nonlinear effects of ISR and the *sensor-to-shooter link* in combat as modeled in SEAS (Frisco, 1999).

### Purpose

The research presented here is intended as a first step in validating object-scaling techniques. This paper will illustrate a validation method suitable for evaluating current object-scaling techniques. The validation method presented will give researchers a powerful tool for validating good – and repairing flawed – techniques.

Significance

It is essential that analysts are able to show how, when, and why a particular scaling technique is used. Conducting analysis while lacking proper connection with theoretical guidelines results in questionable conclusions that are difficult to explain. Current scaling methods involve symmetric or asymmetric vertical slices. In particular, the analyst may scale certain objects by one factor and other objects by a different factor in order to obtain results that *feel* right. (We saw an asymmetric vertical slice when the complex dyke example was scaled.) While the methods by which analysts choose such asymmetric object scaling schemes may be well founded, no prior research exists on which to assert the propriety of their methods.

This effort will advance the analytic community's understanding of object-scaling in computer modeling. It will establish a link between theory and application of the vertical slice methodology. It will present analysts with a straightforward procedure to validate emerging concepts before applying them. This procedure may be employed in a piece-wise fashion – testing only the questionable methods within a particular scenario. In this way, analysts may avoid misdirected analysis and misleading presentation that could cost the Air Force gravely in terms of time and money wasted pursuing the wrong path toward Space Force modernization.

Scope

This initial investigation will lay the groundwork for SEAS validation. Scaling theory application has not been validated with respect to its ability to preserve the

characteristics of the region of interest. A necessary foundation for validation of SEAS is the validation of the use of the *vertical slice methodology* on which SEAS is based.

As an initial exploration, this effort will be well focused. A relatively simple, largely object-independent scenario is used as compared to those used in SEAS-based analysis. The variability of outcomes within each design point is kept low. The investigation is limited to applying scaling to physical objects on the battlefield – e.g. numbers of tanks, aircraft and weapons – and the space in which they operate. It will demonstrate a robust validation method that is not limited by the characteristics of the experimental region.

## II. Related Literature

The analytic community's understanding of non-linear interactions between objects in complex systems is still emerging. Up to this point, much of the development of this paper is based upon unpublished expert opinion and conjecture of those involved with the development of object scaling techniques. The author learned much about vertical aggregation from Capt. Eric Frisco, SMC/XRI, Dr. Robert H. Weber, The Aerospace Corporation, and through the work of Dr. Louis Moore, Rand Corporation.

A vertical slice study conducted by Dr. Moore is discussed below. It is followed by a review of important concepts relative to the problem presented here. The discussion is intended as a review in the areas of simulation output analysis, error in statistical tests, multiple comparison with a control, and response surface metamodels.

### SEAS Vertical Slice Study

Dr. Louis Moore of RAND previously conducted a symmetric object-scaling study in SEAS using a simple case of 9 ground attack aircraft versus 243 armored vehicles at full, one-third, one-ninth, and approximately one-eighteenth scale. The study focused on three points:

1.) The percentage of vehicles killed.
2.) The run-time speed advantage gained by scaling.
3.) The requirements for mitigating the scaling-induced increase in variance.

The study found that the average percentage of vehicles killed was only affected in the smallest slice. Dr. Moore also found the standard deviation decreased as the number of objects increased. Finally, he shows that the ninth-scale scenario is the best

choice for accuracy and efficiency. The study draws no conclusions about the aptness of the ninth-scale slice over any range of input parameter values (1999).

Comparison of Simulation Output

Much of the work on comparing simulation output is limited to comparing output from:

1.) Two different models at the same single design point.
2.) Several different models at the same single design point.
3.) The same model at different design points.
4.) The same model at several different design points.
5.) Two different models at several different design points.

None of these applies in comparing two models over the entire design region. While one might apply the Bonferroni approach to multiple comparisons to cover the design region, it gives no information about the trends occurring between the design points (Banks et al., 1996:475-97; Law and Kelton, 1991:582-603).

Errors in Statistical Tests

Statistical texts define two different errors that can be committed when conducting a statistical test. The first error – a Type I error – is made when a true hypothesis is rejected. The second error is made when a false hypothesis is accepted. This is referred to as a Type II error (Wackerly et al., 1996: 413).

Comparison with a Control

An important and controversial issue when making comparisons with a control involves controlling Type I errors. Some statisticians feel it is not necessary to adjust for dependence among comparisons in a single experiment. Analysts from the simulation camp are no doubt familiar with familywise error rate (FWE) and per comparison error rate (PCE). Recall that FWE is the probability of at least one Type I error and PCE is the expected proportion of Type I errors. Now consider that neither of these takes into account the dependence among multiple comparisons with one control – e.g. a single experiment (Proschan and Follman 1995).

If the *single* experiment for the control goes awry – because of a coding error, a bad random seed, or whatever – then both comparisons in the experiment will be biased. Thus, we find a higher conditional probability of making an error in the second comparison given a mistake was made in the first comparison and so on. In this light, it appears wise to use an independent control for each comparison.

When more than two comparisons are made in a single experiment, Proschan and Follman (1995) show that the number of errors is more variable. The expected number of errors is the same as with comparisons made under separate experiments, but the variance is larger for the single experiment comparisons. Thus, extreme results are more likely under the dependent situation (Proschan and Follman, 1995:4).

Luckily, we may minimize the difference in the way the numbers of errors are distributed between these situations by choosing $k$, the sample size of the control, properly. Proschan and Follman site Dunnett (1955), who "...has shown that power is maximized when the ratio of the control group sample size to the sample size in each

other group is approximately square root of $k$." Therefore, we want a larger sample size for our control group. This is intuitive, since a larger sample-size yields a better estimate. Now, when making two comparisons in a single experiment and using the best sample size for the control, a single control can be used with little difference in the expected number of errors.

Errors in Multiple Comparisons

Sato (1996) gives a historical account of the developing treatment of errors in statistical analysis. A survey of literature on the subject of multiple comparisons reveals an emphasis on controlling Type I error rates. Sato points out that such an emphasis increases the risk of missing significant effects. Hochberg and Tamhane (1987) discuss several examples of comparisons with regard to the type of error to control. They note that in the case of comparing noncompeting treatments with a control, one should be concerned with the PCE. This follows despite the statistical dependence among comparisons since the interpretation of either comparison is unaffected by the other.

Response Surface Metamodels

It is important to recognize the limits associated with the use of metamodels. Law and Kelton treat metamodels with a certain amount of skepticism. They produced a good surface depiction of their example simulation using 420 equally spaced points on a grid, compared with an invalid metamodel derived from a four point designed experiment. They eventually show the reader a compromising 36-point grid that gives a balance between an invalid metamodel and a *super-valid* metamodel derived from an inefficient

design (1991:679-89). A complex scenario may require a complicated design to derive a valid metamodel efficiently. Many efficient designs, which produce good metamodels, can be found in the statistical literature. Choosing and implementing a design depend upon the purpose of the metamodel and the region being studied. For additional treatment of this subject, see Design Considerations on page 28. The reader is also referred to Neter, et al. 1996:1284; Kleijnen, 1987:147-50,312-37 and Myers and Montgomery, 1995:284-7,306-11,314 for discussions on rotatability; orthogonality; D-optimality; design region; and design choice.

## III. Methodology

The method applied in this research assumes that a progression of simplifying assumptions occurs after which the analyst is left with the model to be used in a study. The progression begins with a true system – as it exists in reality or as it is imagined to be in the future. Next, simplifying assumptions allow the system to be modeled – call this the *full-scale model* or *model truth*, though this model is not intended for coding. Finally, the full-scale model is reduced using scaling techniques such as the vertical slice methodology to permit it to be coded and run.

### Analysis Using Computer Models

In studying a complex system using a computer model, analysts may vary one or more of the system parameters and observe the effect of those changes on the output. Various techniques allow the analyst to affect a mapping of the solution space. Such a mapping may provide significant insight to the decision-maker only to the degree that the mapped solution space agrees with the true solution space. Unwanted disagreement – aside from that accounted for and understood – necessarily implies a flawed mapping. Any conclusions drawn from analyzing a flawed surface will themselves be flawed.

### Model Truth as the Control

When modeling future systems, whether any difference from reality exists is necessarily unknowable. Therefore, comparison of the model as coded to reality will be impossible. Instead, assume – as usual – that the difference from reality is accounted for

in reported assumptions and is *well understood*. Then the greatest truth at the analyst's disposal is *model truth* – the true scenario given simplifying assumptions, *prior* to reduction via any postulated scaling technique. Model truth is defined by the full-scale model – an overwhelming model in terms of numbers of objects and interactions.

The output from this *full-scale* model is the nearest thing to reality that is controllable and testable. Thus, the full-scale model will serve as the control – the basis for comparison with all models that claim to parallel it.

## The Nature of the Study

This study is confirmatory in nature. A specific scaling technique has been proposed. An experiment must be conducted to test each technique. When a flawed scaling scheme is *accepted*, a Type II error is committed. With the method presented here and in the context of a confirmatory study, it is more important that few Type II errors are made. If – instead – the analyst wishes to conduct exploratory analysis with this technique, then more Type II errors should be allowed. The newly discovered scaling schemes found would then be analyzed further (Hochberg, Tamhane, 1987:6).

## Deriving the Map

In a given validation study it is necessary, but may not be sufficient to perform point-wise comparison. Such a comparison may show a significant difference between a control and a scaled model, compelling a researcher to incorrectly abandon the postulated scaling technique. Unfortunately, a rejection so made will end the research process without having afforded the researcher every potential insight. Significant insight is

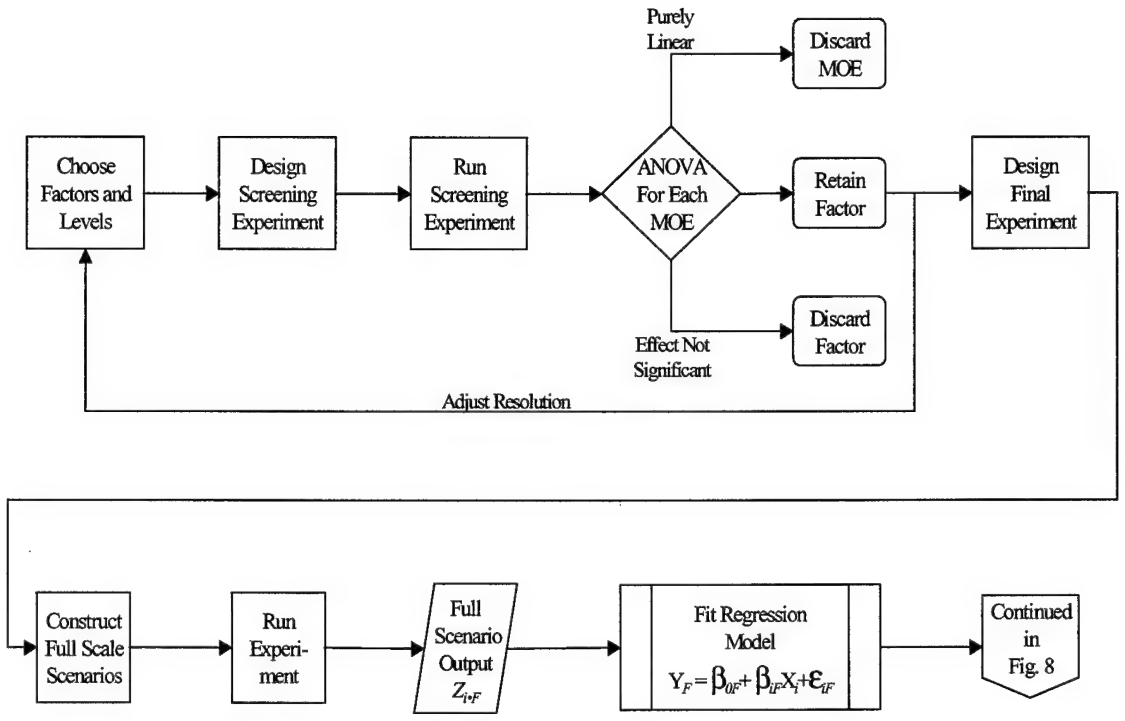


**Figure 7 Fitting a Metamodel**

readily available via comparison of the response surfaces of the models in question. Thus, a surface – or at least a curve – will be mapped for a control and each proposed technique.

Simulation output is generated at each point in a designed experiment (refer to Figure 7 Fitting a Metamodel). The method of least squares is performed to derive a regression function. The input parameters are the same in both the regression and the simulation. The mapping is considered a metamodel – a mathematical representation of the simulation input-output transformation. Now, by performing analysis on the regression function we may analyze the entire region of interest (Kleijnen, 1987:147-50).

Regression Model Assumptions

Before inferences can be drawn from a regression model, the following assumptions must be satisfied:

1.) Collinearity does not exist among the columns of **X**.
2.) Linear regression is applicable.
3.) The simulation responses have constant variance.
4.) The simulation responses are independent.
5.) The simulation responses are normally distributed.
6.) The regression model is correctly specified.

*Collinearity of **X***

Experimental design in simulation studies allows the researcher complete control over the independent variables. Consequently, the columns of **X** will be independent.

*Linearity of Regression Parameters*

The regression function should be linear in $\beta$. Linearity in the regression parameters may be preexisting. Otherwise, a transformation on **X** may remedy the inadequacy. Residual analysis will reveal the need for corrective measures.

*Constant Variance among Simulation Responses*

It may not be clear at the outset whether the constant variance assumption will be satisfied. Kleijnen argues that it is not realistic to assume stochastic simulations will exhibit constant variance among responses. However, the results presented below will show that the dissimilitude in the sample variances of this study were far below Kleijnen's predicted factor of 100 or more. In the event that constant variance is not safely satisfied, the analyst may give more weight to the more reliable observations and

vice-versa. Kleijnen shows that weighted least squares gives an unbiased estimator for β, with the inverse of the $i^{th}$ estimated response variance as the $i^{th}$ weight (1987:162-6).

An important caveat exists for Kleijnen's transformation in the context of comparing metamodels. If one can safely assume that the variance is similarly distributed from the control to the proposed model, then the transformation is safe to use. Otherwise, a different transformation will be performed on each set of responses. The resulting metamodels will not be suitable for comparison, since the relationships among them will have been distorted. Alternatively, the number of replications should be adjusted until the assumption of constant variance is satisfied.

*Simulation Responses Independent and Normally Distributed*

This study uses random seeds to obtain independent responses. The responses must be checked for normality prior to drawing inferences from the regression function. The normality assumption should be assessed during residual analysis.

*Regression Model Specification*

The regression model must be correctly specified. A priori knowledge of the shape of the solution space is helpful. Alternatively, a few well-chosen screening runs will reveal enough important features of the space to suggest the proper regression model. After the model is fit, the analyst must assess its adequacy. Classical statistical techniques may be used in addition to those developed specifically for metamodeling.

Kleijnen suggests the following z* statistic for validating the regression model by comparison with the standard normal variable.

$$z_i^* = \frac{(y_i - \hat{y}_i)}{\sqrt{\hat{\sigma}^2(y_i) + \hat{\sigma}^2(\hat{y}_i)}} \quad \textbf{(1)}$$

Where $i = 1 \ldots n$. The numerator is the difference between the predicted response (using the regression model) and the actual simulation response. The denominator is the estimated standard error for the difference in population means. The comparison may be made at each design point by leaving the point out during the least squares estimation – the $i^{th}$ row vector $x_i$ is deleted from the design matrix (1987:187-9). The Bonferroni inequality gives a lower bound for the family confidence coefficient – the probability that all inferences are correct simultaneously (Neter et al., 1996:153-5).

Typically, the voluminous output from many simulation iterations will mask outlying observations. Fitting the regression model to the average response at each design point will eliminate *masked-outliers*. Kleijnen shows that the resulting regression model is identical when the average response is weighted by its number of replications (1987:195). The law of large numbers guarantees stochastic simulations will produce outlying observations when sufficient replications are made. This technique will help the analyst find true outliers – caused by coding errors and so on – quickly.

Comparing the Solution Space

The procedure that this study follows is shown in Figure 8. Point-wise comparison is conducted on experimental output according to Welch (1938) with the

addition of FWE control.

$$H_0: Z_{i\,F} - Z_{j\,S} = 0, \qquad \forall\, i = j \tag{2}$$

Where $Z_{i\,F}$ is the average loss ratio at point $i$ from the full-scale experiment and $Z_{j\,S}$ is from the scaled-down experiment. A large number of simultaneous null conclusions at this point may compel the researcher to accept the scaling scheme and

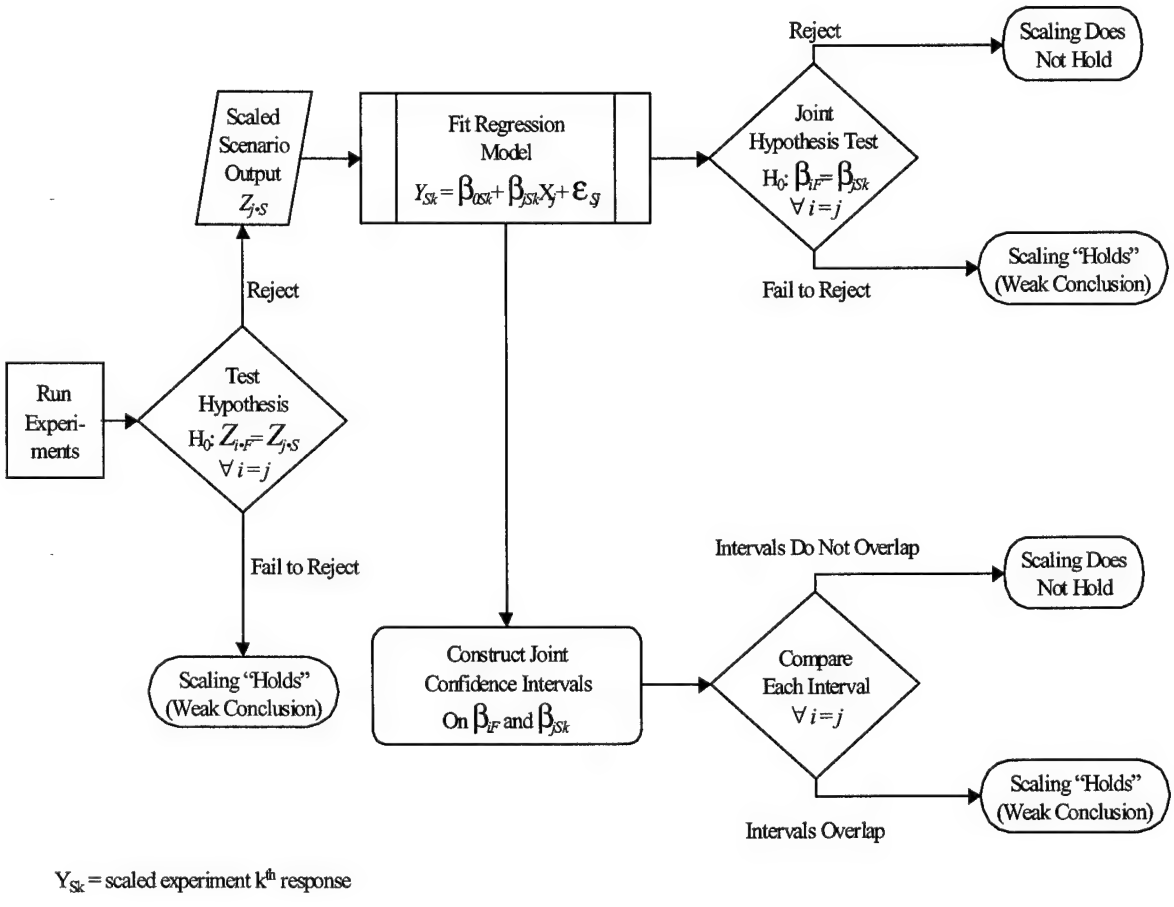


**Figure 8 Statistical Validation of Scaled Scenarios**

suspend further analysis. Conversely, statistically significant differences between the experiments may not be sufficient to abandon the proposed scaling scheme. Now a comparison of the response surfaces may lend significant insight in determining the suitability of the scaled scenarios to represent model-truth. Thus, the analyst maps the solution space of the control.

From the section Regression Model Specification, recall that an adequate model must be fit. In this case, a quadratic model is proposed in two predictor variables:

$$\boldsymbol{Y}_{iF} = \beta_{0F} + \beta_{1F}X_1 + \beta_{2F}X_2 + \beta_{3F}X^2_1 + \beta_{4F}X^2_2 + \beta_{5F}X_1X_2 + \varepsilon_{iF} \qquad (3)$$

Where $\boldsymbol{Y}_{iF}$ is the $i^{th}$ response of the full-scale simulation.

Next, the solution space of the scaled model is mapped, giving:

$$\boldsymbol{Y}_{jSk} = \beta_{0Sk} + \beta_{1Sk}X_1 + \beta_{2Sk}X_2 + \beta_{3Sk}X^2_1 + \beta_{4Sk}X^2_2 + \beta_{5Sk}X_1X_2 + \varepsilon_{jSk} \qquad (4)$$

Where $\boldsymbol{Y}_{jSk}$ is the $j^{th}$ response of the $k^{th}$ scaled simulation.

Finally, poor comparison between (3) and (4) will show that the proposed model yields output inconsistent with that of the control. Conversely, good comparison will fail to prove that the output is inconsistent.

Thus, the point is to show whether the solution space being analyzed in the scaled scenario is statistically different from that of the full-scale scenario. If the method by which analysts build a vertical slice is flawed, then the true solution space is prone to distortion. If the method is sound, then the solution space should not appear different.

Consider the response surfaces pictured in Figure 9 and Figure 10. They are defined by the models in (3) and (4) above. They follow similar trends, but it is clear that they differ from each other. Nevertheless, the difference may not be sufficient to

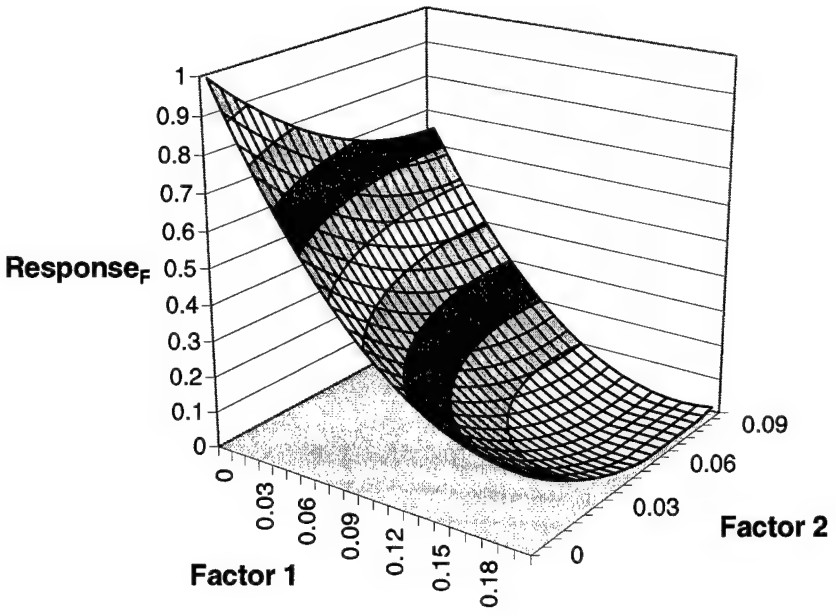


**Figure 9 Notional Surface for $\hat{Y}_F$**

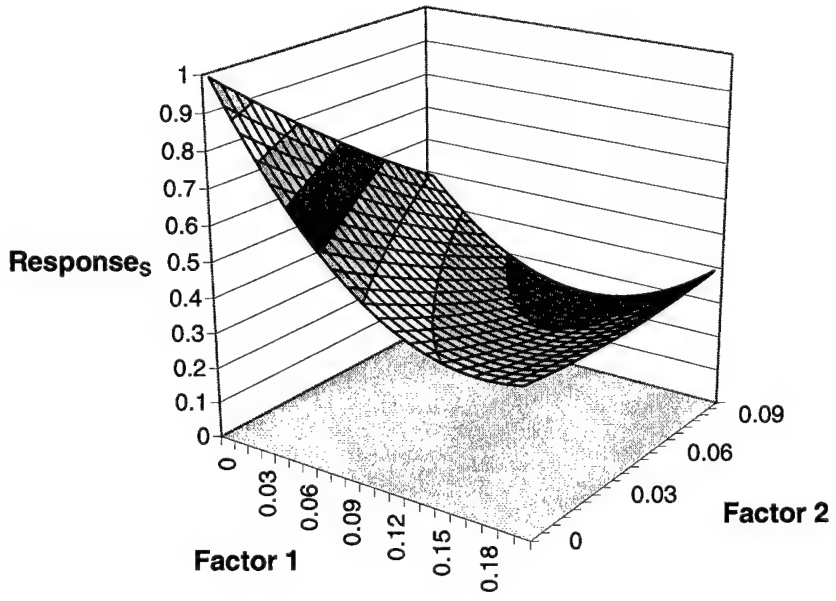


**Figure 10 Notional Surface for $\hat{Y}_S$**

preclude the use of $\hat{Y}_S$ to represent $\hat{Y}_F$. In addition, the surface comparison may lend sufficient insight to suggest how to implement remedial measures on the scaled scenario.

Comparison of simulation output is well represented in the literature and in texts on the subjects of simulation and statistical analysis. The author favors methods that convey insight to those that merely reject a hypothesis. Such insight will allow the analyst to choose an appropriate confidence level for a particular analysis. In this way, analysts may choose a particular scaling method based on its relative merit.

Joint Confidence Intervals

Comparison is made by forming joint confidence intervals on $b_{iF}$ - $b_{jSk}$ – the difference between each regression function coefficient in the control and the scaled model:

$$(\hat{b}_{iF} - \hat{b}_{iS}) \pm t_{(1-\alpha/2g;n_F+n_S-p)} * \hat{\sigma}(b_{iF}, b_{iS}) \qquad \textbf{(5)}$$

Where $g$ is the number of comparisons made, $n$ is the number of responses and $p$ is the number of parameters being compared. Thus, the comparison is between $b_{0F}$ and $b_{0Sk}$, $b_{1F}$ and $b_{1Sk}$, and so on up to $b_{5F}$ and $b_{5Sk}$ for the quadratic model. Note that in the example presented below, noncompeting scaling methods are considered exclusively for comparison with the control. Thus, in the case of the quadratic model in (3) and (4) with no adjustment for PFE, $g = 6$ and $p = 12$. If any one of these intervals fails to contain zero for a choice of $\alpha^*$ (FWE), then the scaling method may be rejected or it may be

studied further. On the other hand, lack of any significant difference at this point indicates a viable scaling scheme.

Since SEAS is used for preliminary analyses, less robust scaling techniques may be acceptable at times. The analyst may find the *p-value* – the smallest level of significance, $\alpha$, for which the data indicate that each parameter is different (Wackerly et al., 1996: 431). Thus, the *p-value* will be presented so that the reader may determine the significance of each comparison.

Design Considerations

Rotatability is considered as a secondary concern, since the regression model is not intended for predicting new responses. An orthogonal design is desirable to minimize the variance of the regression coefficients. A central composite design with axial points placed at radius $\alpha' = 1.581$ will yield near-orthogonality for the quadratic model. Augmenting this design with a D-Optimal search in consideration of a quadratic model

**Table 1 D-Optimal Design**

| Exp Point | Factor 1 | Factor 2 |
|---|---|---|
| 1 | -1 | -1 |
| 2 | 1 | -1 |
| 3 | -1 | 1 |
| 4 | 1 | 1 |
| 5 | -1.581 | 0 |
| 6 | 1.581 | 0 |
| 7 | 0 | -1.581 |
| 8 | 0 | 1.581 |
| 9 | 0 | 0 |
| 10 | -1 | -1 |
| 11 | 1 | -1 |
| 12 | -0.5 | -0.5 |

will enhance the desired minimum variance of the regression coefficients. If independent, constant variance model errors hold, then maximizing the determinant of **X'X** will minimize the volume of the confidence region of ***b*** (Myers, Montgomery, 1995:284-365). The final design is shown in Table 1.

Scenario

The input parameters for the SEAS scenario are summarized in table format in Appendix A. Blue objects are shown for identically modeled objects. When a difference exists, the objects are contrasted in a single table or the difference is notated. The scenario for experimentation is a homogeneous two-wave tank battle with close air support. Red is pitted against blue.

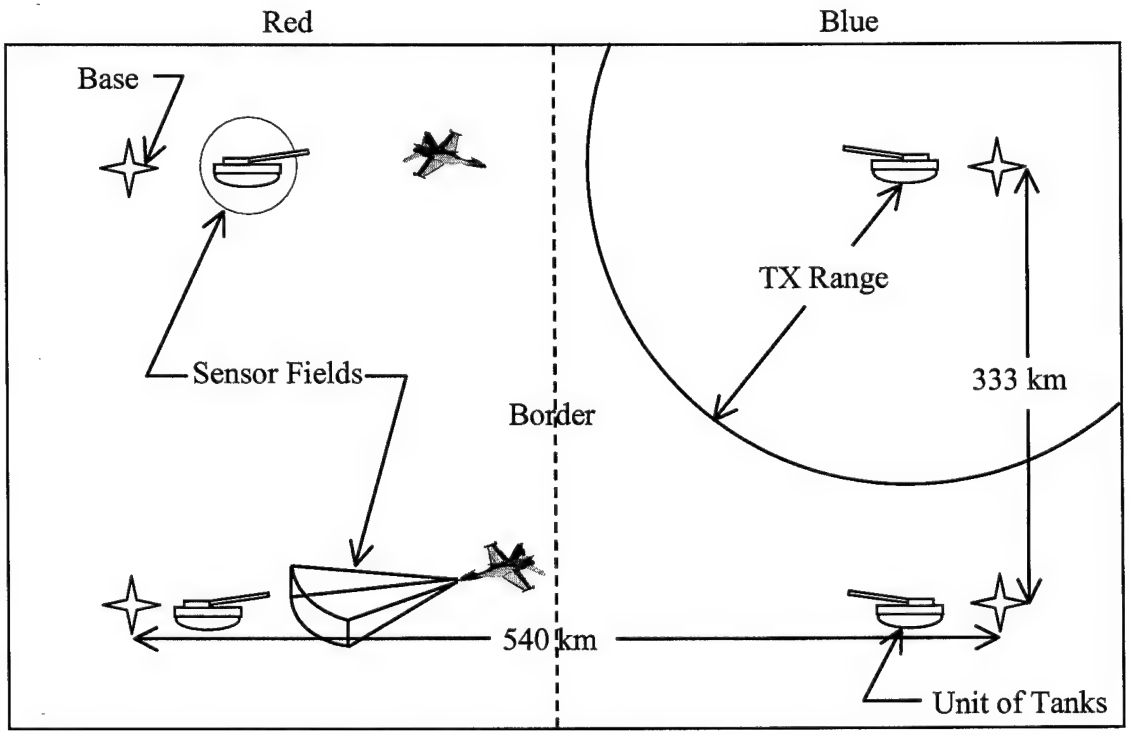


**Figure 11 The Battlefield**

Each side has:

1.) Two hundred generic tanks divided among four units and between two bases.
2.) Forty aircraft divided among four units and between two bases.
3.) Identically scheduled timing events.
4.) Identically listed targets.
5.) Symmetrically located geographic positions.
6.) Identical detection and kill probabilities.
7.) Identical communication capabilities.
8.) Identical sensors.
9.) Identical weapons.

The battlefield is 540 km from red base to blue base and 333 km north to south (refer to Figure 11). Neither weather nor terrain is modeled. Aircraft engagements are not simulated. As tanks from opposing sides approach the border, they are attacked from the air with no means of returning fire.

Aircraft targets are prioritized according to target location. Target lists are shown in Appendix A, Table 7. Notice that aircraft first search for targets at the border. If no targets are found at the border, aircraft will next search for targets at the enemy's base. Aircraft will divert to targets within 250 km as sightings are received. In this scenario, red aircraft may receive such reports from red tanks – not efficient for this task given their short sensor range. Blue aircraft receive sightings from blue tanks and from notional satellite-based sensors with moving target indicators (MTI) – after embellishment.

Excess communication queue capacity exists to avoid lost messages in the full-scale scenario. Aircraft hold target sightings for the expected length of their missions. Aircraft are not allowed to transmit target sightings to each other to help force independence among them. Independence is forced between geographically separated

units to the north and south by forcing objects to traverse the battlefield to the east and west and by limiting aircraft divert distance.

The homogeneous scenario is embellished with notional blue-force space-based ISR assets that will shift the advantage to the blue-force. ISR assets send target location, speed and direction information to blue aircraft. This information cues the aircraft radar to the location of those targets, thus enhancing their probability of detection.

Blue's increased ability causes them to kill reds more quickly. Red's ability is not altered. This necessarily implies blue's losses will diminish with their added ability. The simulation halts when 85% of red forces have been killed. Such a stopping criterion captures the blue advantage at some point during the *heat of battle*. If the simulation drones on too long, one will find red aircraft taking advantage of a target rich environment.

**Table 2 Object Scaling Schemes**

| | BUnit1 | B2Unit1 | BUnit2 | B2Unit2 | BAir1 | B2Air1 | BAir2 | B2Air2 |
|---|---|---|---|---|---|---|---|---|
| Quantity (Full-Scale/ Model Truth) | 50 | 50 | 50 | 50 | 10 | 10 | 10 | 10 |
| Scheme (1) | | | | | | | | |
| Quantity (Half-Scale) | 25 | 25 | 25 | 25 | 5 | 5 | 5 | 5 |
| Quantity (Fifth-Scale) | 10 | 10 | 10 | 10 | 2 | 2 | 2 | 2 |
| Quantity (Tenth-Scale) | 5 | 5 | 5 | 5 | 1 | 1 | 1 | 1 |
| Scheme (2) | | | | | | | | |
| Quantity (Half-Scale) | 50 | 0 | 50 | 0 | 10 | 0 | 10 | 0 |
| Quantity (Fifth-Scale) | 20 | 0 | 20 | 0 | 4 | 0 | 4 | 0 |
| Quantity (Tenth-Scale) | 10 | 0 | 10 | 0 | 2 | 0 | 2 | 0 |

## The Experiments

An identical experiment will be run using models of differing scale. If the theory underlying the vertical slice methodology holds in application, then the output from a well constructed scaled model should necessarily reflect that of the full-scale model (SMC/XR:3; Moore, 1999). Each scaling scheme will have a response surface associated with it. Each surface will be compared to the surface associated with the full-scale scenario. Candidate scaling schemes are listed in Table 2.

## Variables

### *Response*

The blue loss ratio – the number of blue vehicles lost divided by red vehicles lost – is an appropriate Measure of Effectiveness (MOE). The scenario was designed to analyze loss ratio. Further, the loss ratio is expected to show a nonlinear response surface as blue advantage is varied. Studying blue loss ratio also avoids the possibility of dividing by zero.

### *Factors and Levels*

Many SEAS analyses study the effects of varying selected objects' performance parameters such as detection probability (Pd) and cue quality. In this two-factor study, the first factor is per time-increment Pd of a notional space-based sensor with moving target detection capability – MTI Pd. The second factor is the blue aircraft ground radar

cue quality – the benefit derived by the cued sensor. MTI Pd varies from 0.0 to 0.1 – nonexistence to near perfect detection. Cue quality varies from 0.0 to 0.2 – nonexistence to near perfect cueing. Table 17, in Appendix A Scenario Parameters, uses the glimpse model to find the Pd value increase with the time a sensor field covers an object. When cue quality = 0.2 we find that the new Pd ≈ 1.0 after the aircraft's specified 30 minute loiter time (Combat Modeling:4-3).

*Region of Interest and Region of Operability*

Many studies concentrate on the region of interest – the range over which the factors are to be studied. In this notional case, the factors levels are varied over the entire region of operability – the theoretical range over which the independent variables are defined (Myers, Montgomery, 1995:280). This region is easily bounded by the factors in this experiment. Detection probabilities lie in the range (0,1). It follows that the minimum factor levels will yield an expected detection probability of zero and the maximum factor levels will yield an expected detection probability of one. Refer to Appendix A, Table 17. The levels specified in Factors and Levels above will satisfy these conditions in this scenario.

## IV. Results

The experiment described above was run for each of three different scales. Full-scale was run as the control. Referring again to Table 2, the fifth and tenth-scale models were chosen from scheme (1). An initial comparison of the simulation output was made according to equation (2). A large number of simultaneous null conclusions at this point may compel the researcher to accept the scaling scheme and suspend further analysis.

### Point-Wise Comparison

The point-wise comparisons of the difference between the sample means from the experiments with the full and fifth-scale scenarios is shown in Table 3. The comparisons show strong evidence of a difference between the sample means at five of ten experimental points, excluding replicates.

**Table 3 Point-Wise Comparison: Full to Fifth-Scale Experiments**

| Experimental Point | $Z_{i\,F} - Z_{j\,S}$ | *s-pooled* | *Z** | *p-value* | |
|---|---|---|---|---|---|
| 1 | 0.0889 | 0.0059 | 15.0802 | < | 0.0001 |
| 2 | 0.0304 | 0.0053 | 5.6860 | < | 0.0001 |
| 3 | 0.0000 | 0.0054 | 0.0045 | = | 0.9964 |
| 4 | 0.0212 | 0.0048 | 4.3865 | < | 0.0001 |
| 5 | -0.0196 | 0.0081 | 2.4028 | = | 0.0163 |
| 6 | 0.0465 | 0.0047 | 9.8218 | < | 0.0001 |
| 7 | 0.1023 | 0.0059 | 17.2773 | < | 0.0001 |
| 8 | -0.0021 | 0.0053 | 0.4035 | = | 0.6866 |
| 9 | -0.0020 | 0.0055 | 0.3611 | = | 0.7180 |
| 10(r) | 0.0758 | 0.0060 | 12.6351 | < | 0.0001 |
| 11(r) | 0.0395 | 0.0053 | 7.5017 | < | 0.0001 |
| 12 | 0.0006 | 0.0055 | 0.1143 | = | 0.9090 |
| r = replicate | Family confidence coefficient, $p^*$ | | | = | 0.6673 |

Note that in this case and in Table 4 the Bonferroni family confidence coefficient, $p^*$ is a lower bound on the probability that ten inferences in the table – rather than twelve – are correct, simultaneously (Neter et al., 1996:153-5). The reason for the adjustment is the correlation between the pairs one-ten and two-eleven. Identical input parameters between experimental points suggests a high conditional probability that the second hypothesis test will be significant, given the first test was significant and vice-versa.

Table 4 shows strong evidence of a difference between the sample means at six of ten experimental points, again excluding replicates, this time between the full and tenth-scale experiments.

**Table 4 Point-Wise Comparison: Full to Tenth-Scale Experiments**

| Experimental Point | $Z_{i\,F} - Z_{j\,s}$ | *s-pooled* | $z^*$ | *p-value* | |
|---|---|---|---|---|---|
| 1 | 0.0344 | 0.0059 | 5.7958 | < | 0.0001 |
| 2 | -0.0074 | 0.0056 | 1.3228 | = | 0.1859 |
| 3 | -0.0377 | 0.0056 | 6.7351 | < | 0.0001 |
| 4 | -0.0292 | 0.0049 | 5.9363 | < | 0.0001 |
| 5 | 0.0150 | 0.0081 | 1.8493 | = | 0.0644 |
| 6 | -0.0082 | 0.0048 | 1.6924 | = | 0.0906 |
| 7 | 0.0095 | 0.0056 | 1.6927 | = | 0.0905 |
| 8 | 0.1501 | 0.0056 | 26.6061 | < | 0.0001 |
| 9 | -0.0462 | 0.0057 | 8.1292 | < | 0.0001 |
| 10(r) | 0.0314 | 0.0060 | 5.2324 | < | 0.0001 |
| 11(r) | 0.0036 | 0.0054 | 0.6607 | = | 0.5088 |
| 12 | -0.0368 | 0.0056 | 6.6141 | < | 0.0001 |
| r = replicate | Family confidence coefficient, $p^*$ | | | = | 0.9059 |

Since both the fifth and tenth-scale scenarios appear to compare poorly to the control, a metamodel will be derived for each scale. These may lend significant insight in determining the suitability of the scaled scenarios to represent model-truth.

Surface Comparison

The responses were transformed according to the method on page 21, since the assumption of constant variance could not otherwise be satisfied. The number of replicates was adjusted to balance: 1. The estimated variance associated with an estimated mean response with 2. The sample variance associated with the average loss ratio at that point. If not balanced, the subsequent surface comparisons would have misstated any significant difference. Every doubling of the number of replicates will halve the diagonal elements of the variance covariance matrix. Doubling the number of responses will change the mean square error of the regression model. The researcher must discover the best number of replicates by trial and error. Regression models were built using two replicates at each design point. Each replicate represents the first or second half of the replications at a design point (Neter et al., 1996:208-10).

Residual analysis during the model-building process suggested the transformations:

$$X^*_i = e^{-X_i} \tag{6}$$

For $i = 1, 2$.

The estimate for $b_2$ lacked significance in the metamodel for the full-scale experiment. Next, the solution space of each proposed model was mapped. The estimate for $b_2$ lacked significance in the fifth-scale metamodel. Thus, it was dropped from both the full and fifth-scale metamodels. In the tenth-scale metamodel $b_5$ lacked significance. Some will argue that this estimate should be dropped from the tenth-scale metamodel. However, since it is significant in the full-scale model it should remain to allow the

**Table 5 Surface Comparison: Full to Fifth-Scale**

| | | Full-Scale (Control) | | Fifth-Scale | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Est. | Std Error | Est. | Std Error | *bi - bj* | *s-pooled* | *z** | *p-value* |
| Intercept | *b0* | 1.564 | 0.107 | 1.141 | 0.107 | 0.42 | 0.152 | 2.790 | 0.003 |
| e^-X1 | *b1* | 0.502 | 0.047 | 0.512 | 0.047 | -0.01 | 0.067 | 0.146 | 0.442 |
| e^-X1*e^-X1 | *b2* | | | | | | | | |
| e^-X2 | *b3* | 0.009 | 0.099 | 0.188 | 0.100 | -0.18 | 0.141 | 1.270 | 0.102 |
| e^-X2*e^-X2 | *b4* | 0.109 | 0.018 | 0.086 | 0.018 | 0.02 | 0.025 | 0.947 | 0.172 |
| e^-X1*e^-X2 | *b5* | -0.159 | 0.030 | -0.116 | 0.030 | -0.04 | 0.042 | 1.011 | 0.156 |
| | | | | Family confidence coefficient, *p** | | | | | 0.825 |

comparison to be made. The parameter estimates and comparisons are shown in Table 5 and Table 6. See Figure 12 to Figure 14 for surface plots of the metamodels.

From Table 5 notice the significant difference between the intercepts of the full and fifth scale metamodels. Conversely, the estimates for $b_1$ are not significantly different. The remaining comparisons are more difficult to judge. Still, interpretation of the surface comparisons may lead to refinement of the proposed scaling technique that currently appears promising.

The tenth-scale metamodel in Table 6, however, appears decidedly different from the control. The tenth-scale simulation model would likely be dropped, as it does not appear to be a good substitute for model-truth.

**Table 6 Surface Comparison: Full to Tenth-Scale**

| | | Full-Scale (Control) | | Tenth-Scale | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Est. | Std Error | Est. | Std Error | *bi - bj* | *s-pooled* | *z** | *p-value* |
| Intercept | *b0* | 1.564 | 0.107 | 1.840 | 0.105 | -0.28 | 0.150 | 1.839 | 0.033 |
| e^-X1 | *b1* | 0.502 | 0.047 | 0.323 | 0.046 | 0.18 | 0.066 | 2.710 | 0.003 |
| e^-X1*e^-X1 | *b2* | | | | | | | | |
| e^-X2 | *b3* | 0.009 | 0.099 | -0.086 | 0.098 | 0.09 | 0.139 | 0.680 | 0.248 |
| e^-X2*e^-X2 | *b4* | 0.109 | 0.018 | 0.139 | 0.017 | -0.03 | 0.025 | 1.213 | 0.113 |
| e^-X1*e^-X2 | *b5* | -0.159 | 0.030 | -0.044 | 0.029 | -0.11 | 0.042 | 2.748 | 0.003 |
| | | | | Family confidence coefficient, *p** | | | | | 0.920 |

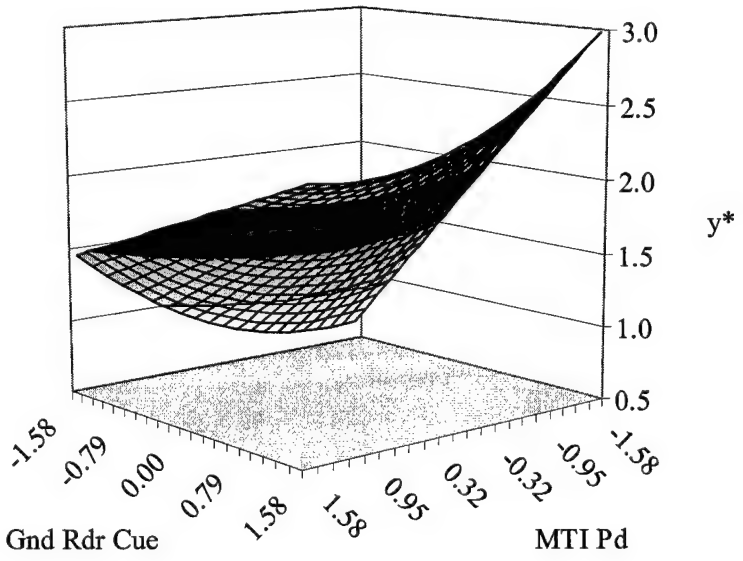


**Figure 12 Full-Scale Surface: Transformed Response**

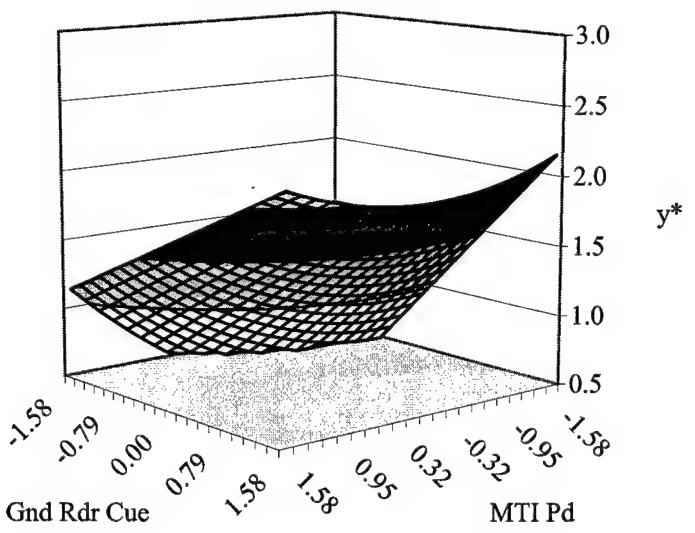


**Figure 13 Fifth-Scale Surface: Transformed Response**

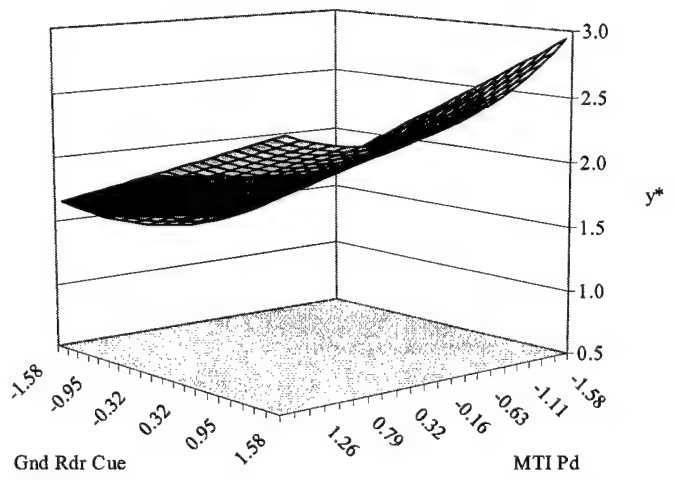


**Figure 14 Tenth-Scale Surface: Transformed Response**

# V. Conclusion

## Challenges and Recommendations

### *Processor Time*

Obviously, the full-scale model will run slowly on today's average personal computer – this is the motivation behind vertical aggregation. The validation approach at the center of this research requires model-truth as the control. The analyst should have more confidence in the control than the hypothesized scaling schemes. This turned out to be problematic for the example used in this paper. The control in the example did not exhibit reduced variance such as that experienced in Dr. Louis Moore's experiment (1999:3). Running additional replications – the preferred remedy – was prohibitive in this case. Unfortunately, an additional 2000 hours of processor time was indicated. Computer time will likely be the most critical resource in this type of research.

If it is not possible to get sufficiently low variance output from the full-scale model in a reasonable number of replications, one should consider using an alternative-scale model for the control. While the full-scale scenario is clearly the best choice for the control – since it is by definition, model-truth – the researcher may find it necessary to use some less prohibitive representation. The particular scale used should obviously be as close to full-scale as the resources can bear.

*Adjusting for Non-Constant Variance*

It is not likely that the responses from the simulation model will exhibit constant variance within a single experiment or from one to the next. It is highly desirable to have the most accurate estimates of the response with minimum constant variance across all experimental points. The transformation outlined in the section Constant Variance among Simulation Responses on page 21 results in a good metamodel for each data set. Unfortunately, it would result in a slightly different transformation for each model based on each simulation's own variance. If the distribution of variance from one model to the next is about the same, then a valid comparison will result. Otherwise, this transformation should be avoided. Here, the variance from the full-scale experiment was used in all three transformations. A better solution will find the number of replications at each point needed to obtain constant variance on all points from all experiments. The extra time it takes will likely pay high dividends in the certainty it lends to the conclusions of the analysis.

*Constant Model Specification*

A single regression model may not be appropriate for all comparisons. It may be difficult to find a model that is valid for all data sets. Compromise may be needed – e.g. a predictor variable with little effect in a scaled model may be included in both models being compared. If, instead, the weak predictor was left out, one model may turn out to be sorely misspecified, preventing any inferences from being drawn. The best solution will be to use the most concise model that is valid for the control. Model building skill will likely be the second critical resource in this type of research.

*Validation in Practice*

The approach presented here is not applicable in validating a SEAS scenario in its entirety. Instead, a modular approach is recommended. The researcher who derives and uses vertical aggregation techniques likely has a greater need to validate one technique over another. Perhaps one technique has been previously validated. Still another enjoys expert approval – validation by consensus. These tried-and-true techniques do not need further validation. However, some of the more aggressive techniques – the more difficult to grasp or explain – are perfect candidates for such an involved method.

*Verifying Input Files*

Since the process is not automated, its execution is arduous. Each experimental point must have its own input file (war file). Each input file – 36 in this example – must be verified for accuracy. Making changes to separate war files invites coding errors that may corrupt the analysis. One method to prevent such errors uses one column in a spreadsheet for each war file so that a text string comparison function can be used to verify them all simultaneously. A macro can then be used to copy and save each file after changes are made. It may be worthwhile to build such functionality directly into the SEAS program.

Summary

The responses among the experiments in this example appear to share great similarities. From Figure 15 alone (Appendix B), it seems that the output is consistent among the different scale models. In the point-wise comparison, we found statistically

significant differences between many pairs of output. Practical significance is left to the discretion of the reader. Finally, in the surface comparison, some bias exists between the surface of the control and the postulated models. The high variance in a few points in the experimental region (Figure 16) makes the power of the surface comparisons difficult to judge. The comparisons themselves can be no more certain than the control.

The method presented used one control by which to judge the merit of two separate simulation models without adjusting for dependence. This was acceptable since Type II errors – made less often – were of primary concern. In practice, the researcher should use independent controls or adjust sample size as outlined in Related Literature (Chapter 2) or as indicated by the estimated variance of the control. In simulation experiments, most would likely prefer combining replications for accurate estimation to separating them into independent, less accurate controls. In that case, a larger sample size will be required for the control.

### Further Research

Prior to proceeding with research based on the example presented here, the replications must be adjusted among the design points to satisfy constant variance. The power of the current surface comparison may then be greater than that indicated in Table 5. A finer resolution experiment may be conducted on the interesting regions of the space. In some regions, greater disparity seems to exist between the full and fifth-scale scenarios (Figure 15). It may be possible to discover conditions under which this scaling scheme is more or less robust.

Once constant variance is met, the second scaling scheme in Table 2 may be investigated to study the effects of scaling surface coverage of the battle. This scenario may also be useful to test the idea of under-scaling aircraft and decreasing weapons load or over-scaling aircraft and decreasing turnaround time. Independence may be removed allowing communication among aircraft. The communication range of tanks and divert range of aircraft may also be increased to remove geographical independence.

Dependencies may be dealt with one at a time. If each of these may be overcome with some scaling technique, it may eventually be possible to study complex systems modeled with their complexity intact, yet in a scaled scenario.

The real benefit of exploratory research comes from new discoveries. Surface comparisons may yield significant insights. Researchers may use such insights to remedy failed scaling techniques. When combined with current methods, the research presented here gives analysts a powerful tool for scientific discovery.

## VI. Appendix A Scenario Parameters

Blue objects are shown for identically modeled objects. Objects are contrasted in a single table or any difference is notated. For all indicators in the following tables, 1 is true and 0 is false.

**Table 7 Targets**

| Objective | Location | Priority | Weapon |
|---|---|---|---|
| B1 | Border | 1 | JSOW2 |
| B2 | Border | 1 | JSOW2 |
| B1 | Base | 2 | JSOW2 |
| B2 | Base | 2 | JSOW2 |

**Table 8 Communication**

| | BlueCue TD/RD | Blue/Red Air TC/RC | Blue/Red FM TC/RC |
|---|---|---|---|
| Channel | SIGQ | BAIRQ/ RAIRQ | BUHFQ/ RUHFQ |
| Range km | 60000 | 300 | 300 |
| Delay min | 1 | 0 | 0 |
| ProbabilityOK | 1 | 1 | 1 |
| MaxMessages | 30000 | 50000 | 50000 |
| MaxRate msgs/min | 1000 | 1000 | 1000 |
| HoldTime min | 10 | 1 | 1 |
| Modes | TX/RX | TX/RX | TX/RX |
| MessageType | Sitreps | Both | Both |
| Jammer | 0 | 0 | 0 |
| Jams | None | None | None |
| Note: T = transmit, R = receive, D = data, C = communication | | | |

Sensors:

1.) Blue AC GNDRDR detects ground vehicles.
2.) B sensor is used by blue tanks to detect red tank. Each detection is communicated to blue aircraft as if to call in an air-strike.
3.) The Moving Target Indicator (MTI) Space-Based Sensor detects moving targets and cues Blue AC GNDRDR.

**Table 9 Sensors**

| | **Blue AC Ground Radar** | **Blue Sensor** | **MTI (Blue only)** |
|---|---|---|---|
| Min Range km | 0 | 0 | 1050 |
| Max Range km | 80 | 4 | 2500 |
| Degrees | 88 | 120 | 60 |
| Cued (Blue only) | 1 | 0 | 1 |
| Cue Range km (Blue only) | 88 | | 2900 |
| Cue Quality (Blue only) Experimental Range | 0 - 0.2 | | 0 - 0.2 |
| Az Limited | 1 | 1 | 0 |
| Az Width | 160 | 180 | |
| TLE m | 5 | 5 | 200 |
| TVE m/min | 0 | 60 | 50 |
| Prob BDA | | 1 | 0 |
| BDA Time min | | 0 | 0 |
| Active | 1 | 0 | 1 |
| Detects | 0 | 0 | 0 |
| MTI | 0 | 0 | 1 |
| Land | 1 | 1 | 1 |
| Air | 0 | 0 | 1 |
| MaxHops | 8 | 3 | 1 |

**Table 10 Blue Sensor Advantages**

| **Space-Based Embellishments** | **Sensor Cue Embellishments** |
|---|---|
| Moving Target Indicator (MTI) Sensor | MTI cues blue aircraft to location of red vehicles |

**Table 11 Weapons**

| | Gun | JSOW2 |
|---|---|---|
| Range m | 3000 | 75000 |
| Kill Radius m | 7 | 15 |
| CEP m | 3 | 3 |
| Reliability | 0.95 | 1 |
| Rate Of Fire round/min | 2 | 1 |
| Offensive Power | 5 | 5 |
| Use Limit | 50 | 2 |
| Land | 1 | 1 |
| Sea | 0 | 1 |
| Air | 0 | 0 |
| Space | 0 | 0 |
| Radar | 0 | 0 |
| Save | 0 | 0 |
| ReArm | 1 | 0 |
| Prioritize | 0 | 1 |
| Coordinate Fire | 1 | 1 |
| Need Local | 1 | 1 |
| Missile | 0 | 0 |
| Fire While Moving | 0 | 1 |

**Table 12 Vehicles**

| | B1 (tank) | B2 (tank) |
|---|---|---|
| Icon | 1705 | 1705 |
| Altitude | 0 | 0 |
| Speed | 36 | 36 |
| Fuel Capacity | 1800 | 1800 |
| Fuel Use | 20 | 20 |
| Fire Wait | 6 | 6 |
| Land | 1 | 1 |
| Sea | 0 | 0 |
| Air | 0 | 0 |
| Comm | Blue FM RC | Blue FM RC |
| Comm | Blue_Air_TC | Blue_Air_TC |
| Sensor | Bsensor | BSensor |
| Weapon | Gun | Gun |

**Table 13 Aircraft**

| Aircraft | Blue Plane | Red Plane |
|---|---|---|
| Icon | 1701 | 1703 |
| Altitude | 7000 | 7000 |
| Speed | 500 | 500 |
| Fuel Capacity | 45000 | 45000 |
| Fuel Use | 400 | 400 |
| Divert Range | 250 | 250 |
| Loiter | 30 | 30 |
| TurnAround | 240 | 240 |
| Sensor | Blue AC GNDRDR | Red AC GNDRDR |
| Weapon | JSOW2 | JSOW2 |
| Comm | Blue Air RC | Red Air RC |
| Comm | BlueCueRD | |
| ThreatHold | 150 | 150 |

**Table 14 BlueBase1 Units**

| | B unit1 | B unit2 | B Air1 | B Air2 |
|---|---|---|---|---|
| Interval | 10 | 10 | 0 | 0 |
| Comm | Blue FM TC | Blue FM TC | Blue FM TC | Blue FM TC |
| Comm | Blue FM RC | Blue FM RC | Blue FM RC | Blue FM RC |
| Vehicle/Plane | B1 (qty-50) | B2 (qty-50) | Blue Plane (qty-10) | Blue Plane (qty-10) |
| Deploy To | Blue Base1 | Blue Base1 | | |
| Action to Take | Hide 3 | Hide 3 | | |
| Wait Until | Wave1 | Wave2 | Begin Air1 | Begin Air2 |
| Action to Take | Hide 0 | Hide 0 | BAFGA List1 | BAFGA List1 |
| Move To | Border1 | Border1 | | |
| Wait Until | Near end | Near end | | |
| Aim Mode | | | Vel | Vel |
| Identical units are located at BlueBase2 for a total of 200 blue tanks and 40 blue aircraft in the full-scale scenario. | | | | |

**Table 15 Forces**

| | Bforce | RForce |
|---|---|---|
| Stance | Pursue | Pursue |
| AttackRatio | 0.9 | 0.9 |
| WithdrawRatio | 0.9 | 0.9 |
| Interval | 20 | 20 |
| Enemy | Rforce | BForce |
| Unit | BUnit1 | RUnit1 |
| Unit | BUnit2 | RUnit2 |
| Unit | BAir1 | RAir1 |
| Unit | BAir2 | RAir2 |
| Unit | B2Unit1 | R2Unit1 |
| Unit | B2Unit2 | R2Unit2 |
| Unit | B2Air1 | R2Air1 |
| Unit | B2Air2 | R2Air2 |
| DeployTo | BlueBase1 | RedBase1 |
| WaitUntil | Nearend | Nearend |

**Table 16 Detection Probabilities (per time step)**

| | Tank1 | Tank2 |
|---|---|---|
| MTI (Blue Only) | 0.04 | 0.04 |
| AC_GNDRDR | 0.01 | 0.01 |
| Sensor (Tank Gun) | 0.03 | 0.015 |

**Table 17 Blue GNDRDR Pd Increase Due to Cue Quality**

| Original Pd | Cue Quality | New Pd | Number of time tics with object in sensor field | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 0.01 | 0 | 0.01 | 0.096 | 0.105 | 0.114 | 0.122 | 0.131 | 0.14 | 0.149 | 0.16 | 0.17 |
| 0.01 | 0.2 | 0.208 | 0.903 | 0.923 | 0.939 | 0.952 | 0.962 | 0.97 | 0.976 | 0.98 | 0.98 |
| | | | Number of time tics with object in sensor field | | | | | | | | |
| | | | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 0.01 | 0 | 0.01 | 0.174 | 0.182 | 0.19 | 0.198 | 0.206 | 0.214 | 0.222 | 0.23 | 0.24 |
| 0.01 | 0.2 | 0.208 | 0.988 | 0.991 | 0.993 | 0.994 | 0.995 | 0.996 | 0.997 | ~1.0 | ~1.0 |

**Table 18 Kill Probabilities**

| | Tank |
|---|---|
| JSOW2 | 0.5 |
| Gun | 0.3 |

## VII. Appendix B Experiment Output

**Table 19 Average Loss Ratio**

| Replications | 600 | 2000 | 2500 |
|---|---|---|---|
| Experimental Point | Full | Fifth | Tenth |
| 1 | 0.315 | 0.349 | 0.404 |
| 2 | 0.274 | 0.266 | 0.304 |
| 3 | 0.305 | 0.268 | 0.305 |
| 4 | 0.192 | 0.163 | 0.213 |
| 5 | 0.637 | 0.652 | 0.617 |
| 6 | 0.185 | 0.176 | 0.231 |
| 7 | 0.467 | 0.476 | 0.569 |
| 8 | 0.229 | 0.180 | 0.227 |
| 9 | 0.265 | 0.219 | 0.263 |
| 10 | 0.322 | 0.354 | 0.398 |
| 11 | 0.266 | 0.269 | 0.305 |
| 12 | 0.299 | 0.262 | 0.300 |

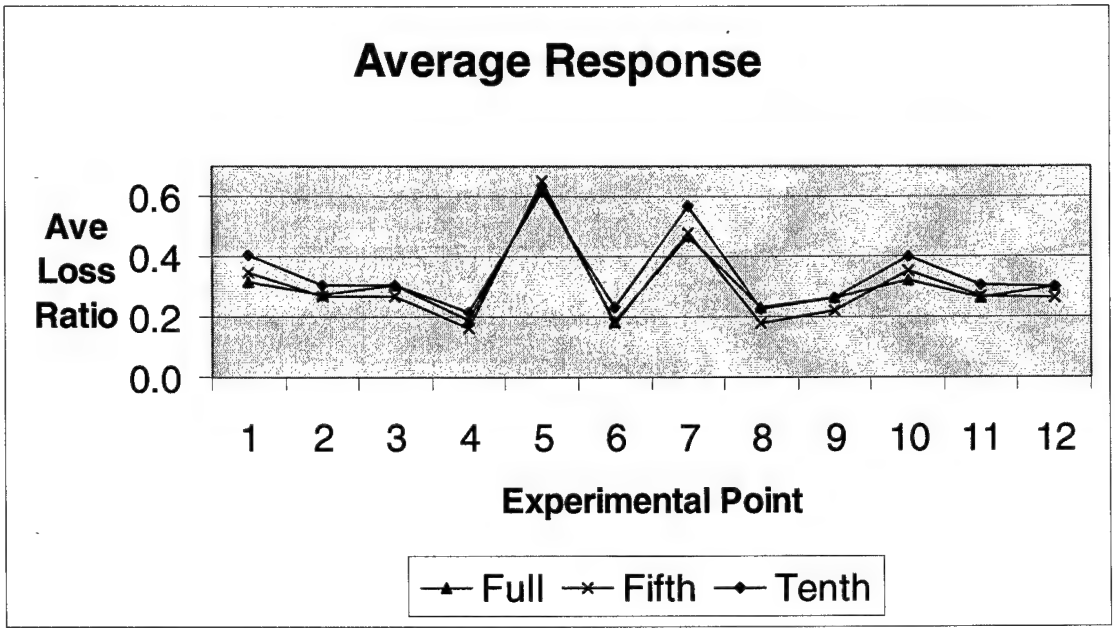


**Figure 15 Average Loss Ratio**

**Table 20 Sample Variance of Average Loss Ratio**

| Replications | 600 | 2000 | 2500 |
|---|---|---|---|
| Experimental Point | Full | Fifth | Tenth |
| 1 | 0.0161 | 0.0166 | 0.0197 |
| 2 | 0.0137 | 0.0166 | 0.0144 |
| 3 | 0.0140 | 0.0159 | 0.0143 |
| 4 | 0.0112 | 0.0111 | 0.0116 |
| 5 | 0.0335 | 0.0200 | 0.0262 |
| 6 | 0.0103 | 0.0124 | 0.0131 |
| 7 | 0.0136 | 0.0179 | 0.0309 |
| 8 | 0.0143 | 0.0118 | 0.0113 |
| 9 | 0.0152 | 0.0140 | 0.0131 |
| 10 | 0.0165 | 0.0172 | 0.0213 |
| 11 | 0.0131 | 0.0149 | 0.0146 |
| 12 | 0.0146 | 0.0135 | 0.0147 |

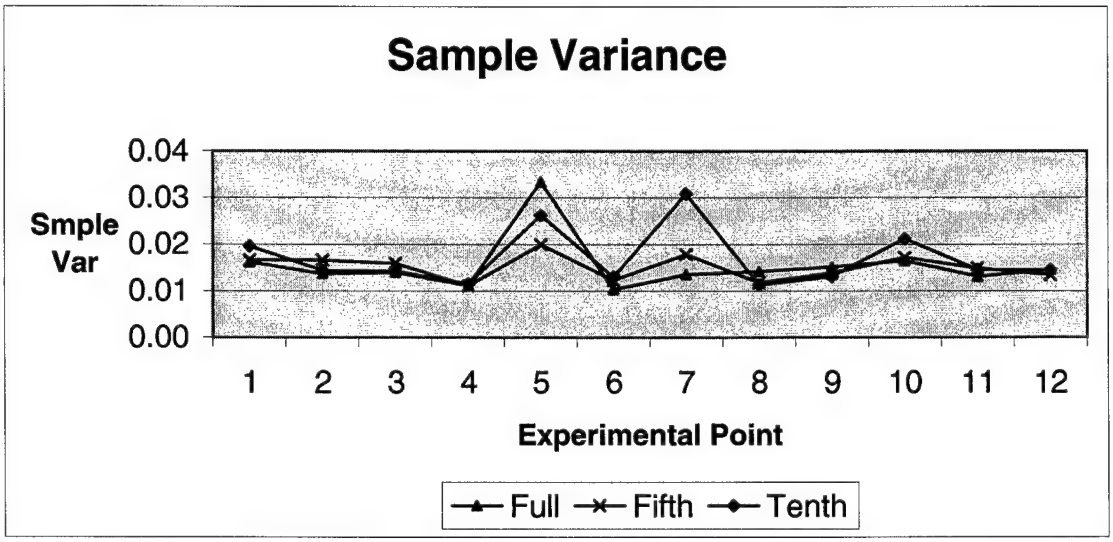


**Figure 16 Sample Variance of Average Loss Ratio**

**Table 21 Average Loss Ratio: 2 Replicates**

| Replications | | 300 | 1000 | 1250 |
|---|---|---|---|---|
| Experimental Point | Replicate | Full | Fifth | Tenth |
| 1 | 1 | 0.316 | 0.349 | 0.406 |
| 1 | 2 | 0.314 | 0.349 | 0.401 |
| 2 | 1 | 0.277 | 0.269 | 0.301 |
| 2 | 2 | 0.270 | 0.263 | 0.307 |
| 3 | 1 | 0.304 | 0.267 | 0.305 |
| 3 | 2 | 0.306 | 0.268 | 0.306 |
| 4 | 1 | 0.194 | 0.165 | 0.211 |
| 4 | 2 | 0.190 | 0.161 | 0.216 |
| 5 | 1 | 0.628 | 0.652 | 0.616 |
| 5 | 2 | 0.646 | 0.652 | 0.619 |
| 6 | 1 | 0.180 | 0.173 | 0.229 |
| 6 | 2 | 0.189 | 0.180 | 0.233 |
| 7 | 1 | 0.472 | 0.475 | 0.573 |
| 7 | 2 | 0.461 | 0.478 | 0.566 |
| 8 | 1 | 0.223 | 0.178 | 0.227 |
| 8 | 2 | 0.235 | 0.183 | 0.227 |
| 9 | 1 | 0.262 | 0.218 | 0.264 |
| 9 | 2 | 0.268 | 0.219 | 0.261 |
| 10 | 1 | 0.325 | 0.355 | 0.393 |
| 10 | 2 | 0.320 | 0.353 | 0.403 |
| 11 | 1 | 0.268 | 0.267 | 0.300 |
| 11 | 2 | 0.264 | 0.272 | 0.310 |
| 12 | 1 | 0.282 | 0.260 | 0.297 |
| 12 | 2 | 0.316 | 0.264 | 0.303 |

**Table 22 Sample Variance of Average Loss Ratio: 2 Replicates**

| Replications | | 300 | 1000 | 1250 |
|---|---|---|---|---|
| Experimental Point | Replicate | Full | Fifth | Tenth |
| 1 | 1 | 0.0162 | 0.0163 | 0.0200 |
| 1 | 2 | 0.0161 | 0.0170 | 0.0195 |
| 2 | 1 | 0.0130 | 0.0163 | 0.0145 |
| 2 | 2 | 0.0144 | 0.0169 | 0.0143 |
| 3 | 1 | 0.0150 | 0.0162 | 0.0143 |
| 3 | 2 | 0.0131 | 0.0157 | 0.0143 |
| 4 | 1 | 0.0118 | 0.0109 | 0.0124 |
| 4 | 2 | 0.0107 | 0.0113 | 0.0108 |
| 5 | 1 | 0.0323 | 0.0193 | 0.0273 |
| 5 | 2 | 0.0346 | 0.0207 | 0.0252 |
| 6 | 1 | 0.0101 | 0.0121 | 0.0125 |
| 6 | 2 | 0.0105 | 0.0126 | 0.0137 |
| 7 | 1 | 0.0139 | 0.0181 | 0.0299 |
| 7 | 2 | 0.0134 | 0.0177 | 0.0318 |
| 8 | 1 | 0.0138 | 0.0110 | 0.0115 |
| 8 | 2 | 0.0147 | 0.0126 | 0.0112 |
| 9 | 1 | 0.0151 | 0.0141 | 0.0128 |
| 9 | 2 | 0.0152 | 0.0138 | 0.0134 |
| 10 | 1 | 0.0166 | 0.0166 | 0.0208 |
| 10 | 2 | 0.0164 | 0.0178 | 0.0218 |
| 11 | 1 | 0.0124 | 0.0148 | 0.0148 |
| 11 | 2 | 0.0139 | 0.0151 | 0.0143 |
| 12 | 1 | 0.0149 | 0.0130 | 0.0151 |
| 12 | 2 | 0.0137 | 0.0140 | 0.0142 |

## VIII. Appendix C Model Specification

### Table 23 Full-Scale ANOVA

| Response: | Y*Full | | | | |
|---|---|---|---|---|---|
| Summary of Fit | | | | | |
| RSquare | 0.94513 | | | | |
| RSquare Adj | 0.93358 | | | | |
| Root Mean Square Error | 0.16520 | | | | |
| Mean of Response | 2.49302 | | | | |
| Observations (or Sum Wgts) | 24 | | | | |
| | | | | | |
| Effect Test | | | | | |
| Source | Nparm | DF | Sum of Squares | F Ratio | Prob>F |
| e^-X1 | 1 | 1 | 3.0835231 | 112.982 | <.0001 |
| e^-X2 | 1 | 1 | 0.0002258 | 0.0083 | 0.9285 |
| e^-X2*e^-X2 | 1 | 1 | 1.0645447 | 39.0057 | <.0001 |
| e^-X1*e^-X2 | 1 | 1 | 0.7769098 | 28.4665 | <.0001 |
| | | | | | |
| Whole-Model Test | | | | | |
| Analysis of Variance | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Ratio | |
| Model | 4 | 8.9326222 | 2.23316 | 81.8244 | |
| Error | 19 | 0.5185489 | 0.02729 | Prob>F | |
| C Total | 23 | 9.4511711 | | <.0001 | |

**Table 24 Fifth-Scale ANOVA**

| Response: | Y*Fifth | | | | |
|---|---|---|---|---|---|
| Summary of Fit | | | | | |
| RSquare | 0.96218 | | | | |
| RSquare Adj | 0.95422 | | | | |
| Root Mean Square Error | 0.16569 | | | | |
| Mean of Response | 2.39983 | | | | |
| Observations (or Sum Wgts) | 24 | | | | |
| | | | | | |
| Effect Test | | | | | |
| Source | Nparm | DF | Sum of Squares | F Ratio | Prob>F |
| e^-X1 | 1 | 1 | 3.2046972 | 116.721 | <.0001 |
| e^-X2 | 1 | 1 | 0.0974114 | 3.5479 | 0.075 |
| e^-X2*e^-X2 | 1 | 1 | 0.6565021 | 23.9111 | 0.0001 |
| e^-X1*e^-X2 | 1 | 1 | 0.4159032 | 15.148 | 0.001 |
| | | | | | |
| Whole-Model Test | | | | | |
| Analysis of Variance | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Ratio | |
| Model | 4 | 13.272271 | 3.31807 | 120.850 | |
| Error | 19 | 0.521663 | 0.02746 | Prob>F | |
| C Total | 23 | 13.793934 | | <.0001 | |

**Table 25 Tenth-Scale ANOVA**

| Response: | Y*Tenth | | | | |
|---|---|---|---|---|---|
| Summary of Fit | | | | | |
| RSquare | 0.965425 | | | | |
| RSquare Adj | 0.958146 | | | | |
| Root Mean Square Error | 0.162138 | | | | |
| Mean of Response | 2.767661 | | | | |
| Observations (or Sum Wgts) | 24 | | | | |
| | | | | | |
| Effect Test | | | | | |
| Source | Nparm | DF | Sum of Squares | F Ratio | Prob>F |
| e^-X1 | 1 | 1 | 1.273715 | 48.4508 | <.0001 |
| e^-X2 | 1 | 1 | 0.020267 | 0.7709 | 0.3909 |
| e^-X2*e^-X2 | 1 | 1 | 1.722568 | 65.5247 | <.0001 |
| e^-X1*e^-X2 | 1 | 1 | 0.060148 | 2.288 | 0.1468 |
| | | | | | |
| Whole-Model Test | | | | | |
| Analysis of Variance | | | | | |
| Source | DF | Sum of Squares | Mean Square | F Ratio | |
| Model | 4 | 13.947 | 3.48675 | 132.6324 | |
| Error | 19 | 0.499488 | 0.02629 | Prob>F | |
| C Total | 23 | 14.44649 | | <.0001 | |

**Table 26 Distribution Full-Scale Residual**

| Quantiles | | | | | |
|---|---|---|---|---|---|
| maximum | 100.00% | 0.42582 | Moments | | |
| | 99.50% | 0.42582 | Mean | 0 | |
| | 97.50% | 0.42582 | Std Dev | 0.15015 | |
| | 90.00% | 0.14543 | Std Error Mean | 0.03065 | |
| quartile | 75.00% | 0.09347 | Upper 95% Mean | 0.0634 | |
| median | 50.00% | 0.02287 | Lower 95% Mean | -0.0634 | |
| quartile | 25.00% | -0.1101 | N | 24 | |
| | 10.00% | -0.1755 | Sum Weights | 24 | |
| | 2.50% | -0.3079 | | | |
| | 0.50% | -0.3079 | Test for Normality | W | Prob<W |
| Minimum | 0.00% | -0.3079 | Shapiro-Wilk W Test | 0.947569 | 0.2465 |

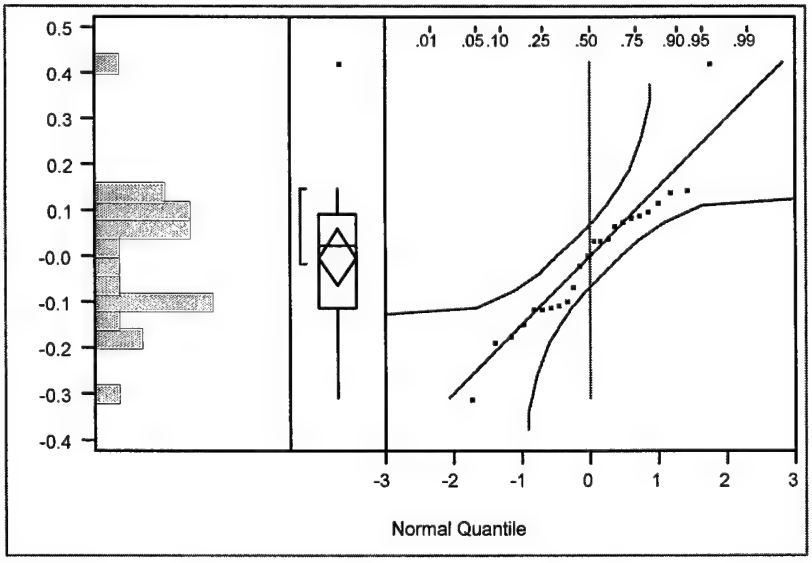


**Figure 17 Normal Probability Full-Scale Residual**

**Table 27 Distribution Fifth-Scale Residual**

| Quantiles | | | | | |
|---|---|---|---|---|---|
| maximum | 100.00% | 0.29024 | Moments | | |
| | 99.50% | 0.29024 | Mean | 0 | |
| | 97.50% | 0.29024 | Std Dev | 0.1506 | |
| | 90.00% | 0.24396 | Std Error Mean | 0.03074 | |
| quartile | 75.00% | 0.12055 | Upper 95% Mean | 0.06359 | |
| median | 50.00% | -0.0365 | Lower 95% Mean | -0.06359 | |
| quartile | 25.00% | -0.0795 | N | 24 | |
| | 10.00% | -0.1629 | Sum Weights | 24 | |
| | 2.50% | -0.3184 | | | |
| | 0.50% | -0.3184 | Test for Normality | W | Prob<W |
| minimum | 0.00% | -0.3184 | Shapiro-Wilk W Test | 0.963516 | 0.518 |

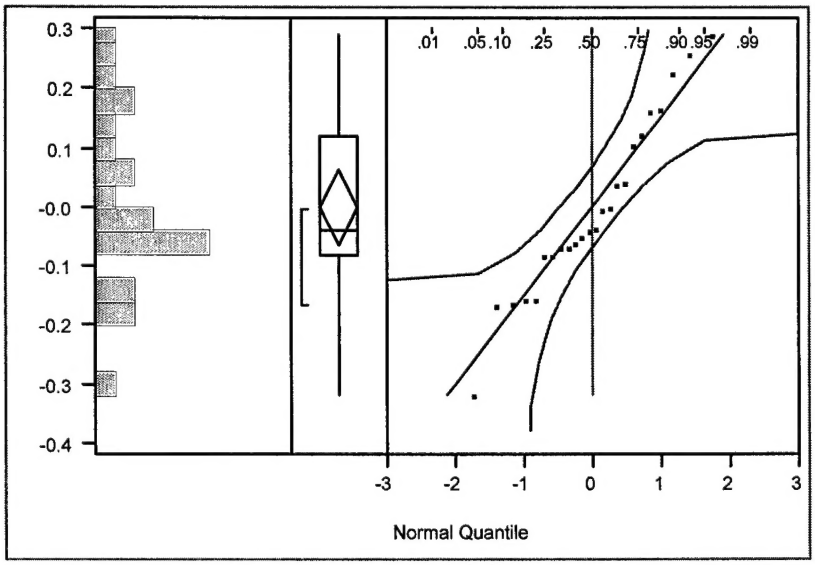


**Figure 18 Normal Probability Fifth-Scale Residual**

**Table 28 Distribution Tenth-Scale Residual**

| Quantiles | | | | | |
|---|---|---|---|---|---|
| maximum | 100.00% | 0.33088 | Moments | | |
| | 99.50% | 0.33088 | Mean | 0 | |
| | 97.50% | 0.33088 | Std Dev | 0.14737 | |
| | 90.00% | 0.25097 | Std Error Mean | 0.03008 | |
| quartile | 75.00% | 0.08056 | Upper 95% Mean | 0.06223 | |
| median | 50.00% | -0.0189 | Lower 95% Mean | -0.06223 | |
| quartile | 25.00% | -0.0771 | N | 24 | |
| | 10.00% | -0.1903 | Sum Weights | 24 | |
| | 2.50% | -0.2716 | | | |
| | 0.50% | -0.2716 | Test for Normality | W | Prob<W |
| minimum | 0.00% | -0.2716 | Shapiro-Wilk W Test | 0.958189 | 0.4102 |

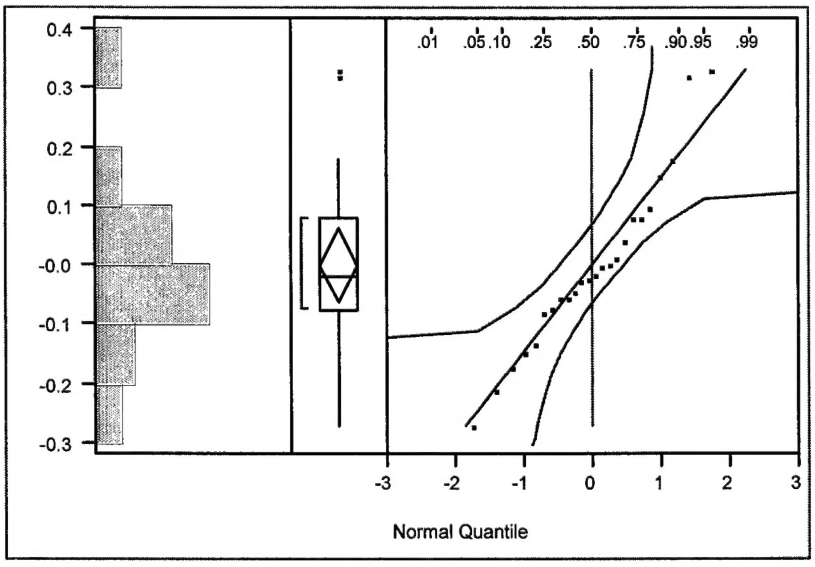


**Figure 19 Normal Probability Tenth-Scale Residual**

## IX. Bibliography

J. Banks, J. S. Carson and B. L. Nelson. Discrete-Event System Simulation. New Jersey: Prentice-Hall, 1996.

Combat Modeling. Unpublished class notes. OPER 671, Combat Modeling I. School of Engineering, Air Force Institute Of Technology (AU), Wright-Patterson Air Force Base, Ohio, October 1999.

T. Czerwinski. Coping with the Bounds: Speculations on Nonlinearity in Military Affairs. Washington, DC: National Defense University. 1998.

P. K. Davis. Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. Santa Monica, CA: Rand Corporation. 1992.

J. A. Dewar, J. J. Gillogy, M. L. Juncosa. Non-Monotonicity, Chaos, and Combat Models. Santa Monica, CA: Rand Corporation. 1991.

C. W. Dunnett. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," Journal of the American Statistical Association, 50: 1096-1121. (1955).

E. W. Frisco. Analyst, US Air Force Space and Missile Systems Center (SMC/XR). Los Angeles AFB, CA. Personal interview. 14 September 1999.

Y. Hochberg and A. C. Tamhane. Multiple Comparison Procedures. New York: John Wiley.& Sons, 1987.

J. P.C. Kleijnen. Statistical Tools For Simulation Practitioners. New York: Marcel Dekker, 1987.

A. M. Law and W. D. Kelton. Simulation Modeling & Analysis. New York: McGraw-Hill Inc., 1991.

L. R. Moore. Untitled draft paper on vertical slice methodology (symmetric scaling) in SEAS. Rand Corporation. 1999.

R. H. Myers and.D. C. Montgomery. Response Surface Methodology. New York: John Wiley & Sons, 1995.

J. Neter, M. H. Kutner, C. J. Nachtsheim.and.W. Wasserman. Applied Linear Statistical Models. Chicago, IL: Richard D. Irwin, Inc., 1996.

T. Sato. "Type I and Type II Error in Multiple Comparisons," Journal of Interdisciplinary & Applied Psychology, 130-3: 292, (May 1996).

M. A. Proschan, D. A. Follman. "Multiple comparisons with control in a single experiment versus separate experiments: Why do we Feel Differently?," American Statistician, 49-2: 144. (May 1995).

M. Shaked. "On Mixtures from Exponential Families," Journal of the Royal Statistical Society, 42B: 192-198. (1980).

T. R Tighe. "Strategic Effects of Airpower and Complex Adaptive Agents: an Initial Investigation," MS Thesis, AFIT/GOA/ENS/99M, Air Force Institute Of Technology (AU), Wright-Patterson Air Force Base, Ohio, March 1999.

US Air Force Space and Missile Systems Center (SMC/XR). System Effectiveness Analysis Simulation Analyst Manual. Los Angeles AFB, CA.

US Air Force Space and Missile Systems Center (SMC/XR). Rules of Thumb, Office Memo on Vertical Slice Methodology. Los Angeles AFB, CA.

D. D.Wackerly, W. Mendenhall, R. L. Scheaffer. Mathematical Statistics with Applications. Belmont, CA: Wadsworth Publishing Company. 1996.

R. H.Weber. The Aerospace Corporation. Los Angeles AFB, CA. Personal interview. September 1999.